Nonparametric Regression Estimation Based on Spatially Inhomogeneous Data: Minimax Global Convergence Rates and Adaptivity

Anestis Antoniadis,
Laboratoire Jean Kuntzmann, Universite Joseph Fourier,
38041 Grenoble CEDEX 9, France.
Email: Anestis.Antoniadis@imag.fr

Marianna Pensky,
Department of Mathematics, University of Central Florida,
Orlando, FL 32816-1364, USA.
Email: Marianna.Pensky@ucf.edu

and

Theofanis Sapatinas,
Department of Mathematics and Statistics, University of Cyprus,
P.O. Box 20537, CY 1678 Nicosia, Cyprus.
Email: fanis@ucy.ac.cy

Abstract

We consider the nonparametric regression estimation problem of recovering an unknown response function f on the basis of spatially inhomogeneous data when the design points follow a known compactly supported density g with a finite number of well separated zeros. In particular, we consider two different cases: when g has zeros of a polynomial order and when g has zeros of an exponential order. These two cases correspond to moderate and severe data losses, respectively. We obtain asymptotic minimax lower bounds for the global risk of an estimator of f and construct adaptive wavelet nonlinear thresholding estimators of f which attain those minimax convergence rates (up to a logarithmic factor in the case of a zero of a polynomial order), over a wide range of Besov balls.

The spatially inhomogeneous ill-posed problem that we investigate is inherently more difficult than spatially homogeneous problems like, e.g., deconvolution. In particular, due to spatial irregularity, assessment of minimax global convergence rates is a much harder task than the derivation of minimax local convergence rates studied recently in the literature. Furthermore, the resulting estimators exhibit very different behavior and minimax global convergence rates in comparison with the solution of spatially homogeneous ill-posed problems. For example, unlike in deconvolution problem, the minimax global convergence rates are greatly influenced not only by the extent of data loss but also by the degree of spatial homogeneity of f. Specifically, even if $1/g$ is not integrable, one can recover f as well as in the case of an equispaced design (in terms of minimax global convergence rates) when it is homogeneous enough since the estimator is "borrowing strength" in the areas where f is adequately sampled.

Keywords: Adaptivity, Besov spaces, inhomogeneous data, minimax estimation, nonparametric regression, thresholding, wavelet estimation.

AMS (2000) Subject Classification: Primary: 62G08, Secondary: 62G05, 62G20

1 Introduction

Applicability of majority of techniques for estimation in the nonparametric regression model rests on the assumption that data is equispaced and complete. These assumptions were mainly adopted by signal processing community where the signal is assumed to be recorded at equal intervals in time. However, in reality, due to unexpected losses of data or limitations of data sampling techniques, data may fail to be equispaced and complete. To this end, we consider the problem of recovering an unknown response function $f \in L^2([0,1])$ on the basis of irregularly spaced observations, i.e., when one observes y_i governed by

$$y_i = f(x_i) + \sigma\, \xi_i, \quad i = 1, 2, \ldots, n, \tag{1.1}$$

where $x_i \in [0,1]$, $i = 1, 2, \ldots, n$, are fixed (non-equidistant) or random points, ξ_i, $i = 1, 2, \ldots, n$, are independent standard Gaussian random variables and $\sigma^2 > 0$ (the noise level) is assumed to be known and finite. Model (1.1) can be viewed as a problem of recovering a signal when part of data is lost (e.g., in cell phone use) or unavailable (e.g., in military applications). Model (1.1) is also intimately connected to the problem of missing data since points x_i, $i = 1, 2, \ldots, n$, can be viewed as the remainder of N equidistant points j/N, $j = 1, 2, \ldots, N$, after observations at $(N-n)$ points have been lost. However, there is a great advantage in treating the missing data problem as a particular case of a nonparametric regression problem: with the last decade seeing tremendous advancement in the field of nonparametric statistics, a nonparametric regression approach to incomplete data brings along all the modern tools in this field such as minimax rates of convergence, Besov spaces, wavelets and adaptive estimators.

The problem of estimating an unknown response function in the context of wavelet thresholding in the nonparametric regression setting with irregular design has been now addressed by many authors, see, e.g., Hall and Turlach (1997), Antoniadis and Pham (1998), Cai and Brown (1998), Sardy *et al.* (1999), Kovac and Silverman (2000), Pensky and Vidakovic (2001), Brown *et al.* (2002), Zhang *et al.* (2002), Kohler (2003) and Amato *et al.* (2006). Several tools were suggested for attacking the problem; here, we shall review only few of them. For instance, the procedure of Kovac and Silverman (2000) relies upon a linear interpolation transformation R to the observed data vector $y = (y_1, y_2, \ldots, y_n)$ that maps it to a new vector of size 2^J $(2^{J-1} < n \leq 2^J)$, corresponding to a new design with equispaced points. After the transformation, the new vector is multivariate normal with mean Rf and covariance matrix which is assumed to have a finite bandwidth, so that the computational complexity of their algorithm is of order n. Cai and Brown (1998) attacked the problem by using multiresolution analysis, projection and wavelet nonlinear thresholding while Sardy *et al.* (1999) applied an isometric method. Pensky and Vidakovic (2001) estimated the conditional expectation $\mathbb{E}(Y|X)$ directly by constructing its wavelet expansion, while Amato *et al.* (2006) applied a reproducing kernel Hilbert space (RKHS) approach in the spirit of Wahba (1990). However, until very recently, all studies have been carried out under the assumption that the nonequispaced design still possesses some regularity, namely, the density function g of the design points x_i, $i = 1, 2, \ldots, n$, is uniformly bounded from below, i.e., $\inf_{x\in[0,1]} g(x) \geq c$ for some constant $c > 0$. In this case, asymptotically, model (1.1) is equivalent to the case of the standard (equispaced) nonparametric regression model, as long as the design density function g is known (see, e.g., Brown *et al.* (2002)).

Recently, an attempt has been made of more advanced investigations of the problem. Kerkyacharian and Picard (2004) introduced warped wavelets to construct estimators of the unknown response function f under model (1.1) when the design density function g has zeros of polynomial order. They, however, measured the error of their suggested estimator in the warped Besov spaces which is, practically, equivalent to measuring the error of the estimator at the design points only. For this reason, the derived estimators posses the usual asymptotical (as the sample size increases)

minimax rates of convergence which do not depend on the order of the zeros of the design density function g. This line of investigation was continued by Chesneau (2007a, 2007b) who constructed asymptotic minimax lower bounds over a wide range of Besov balls, under the assumption that the design density function g is known and that $1/g$ is integrable, and, furthermore, suggested adaptive wavelet thresholding estimators for the unknown response function f. However, in both Kerkyacharian and Picard (2004) and Chesneau (2007a, 2007b), the assumptions on the design density function g are restrictive enough so that the asymptotical minimax rates of convergence of any estimator coincide with the asymptotical minimax rates of convergence under the assumption that g is bounded from below, i.e., the corresponding nonparametric estimation problem is a *well-posed* problem.

Gaïffas (2005, 2006, 2007, 2009) seems to be the only author who considered the problem as an ill-posed problem. He studied local minimax rates and constructed locally adaptive estimators on the basis of local polynomials. Gaïffas (2005, 2007) constructed pointwise estimators of a regression function when $1/g$ is not integrable and showed that convergence rates of the estimators are slower than in the case when g is bounded below, hence, demonstrating that the problem of regression estimation under irregular design is an ill-posed problem. The shortcoming of his work is that the minimax rates of convergence are expressed in a very complex form which is very hard to obtain for a regression function f which belongs to a standard functional class. Also, his techniques are intended for local reconstruction and depend on cross-validation at each point, so that they become too involved when applied to the whole domain of function f.

The objective of the present paper is to study how zeros of the design function g affect reconstruction of regression function f globally. As we show below (see Remark 2), assessing minimax global convergence rates is a much harder task than assessing minimax local convergence rates. Model (1.1) can be viewed as the spatially inhomogeneous ill-posed problem which is inherently more difficult to than spatially homogeneous problems, e.g., deconvolution, especially, the case when the true regression function is spatially homogeneous. To the best of our knowledge, so far, there are no results for asymptotic rates of convergence in the case of spatially inhomogeneous ill-posed problem when its solution is spatially homogeneous since the authors usually avoid the problem by restricting their attention to the case when the estimated function is spatially inhomogeneous, or, at most, belong to a Sobolev ball (see, e.g., Hoffmann and Reiss (2008)).

In what follows, we address these issues. In particular, we mainly consider two different cases: when some fractional power of $1/g$ is integrable (zero of a polynomial order) and when no fractional power of g is integrable (zero of an exponential order). We obtain asymptotic minimax lower bounds for the global risk of an estimator of f and construct adaptive wavelet nonlinear thresholding estimators of f which attain those convergence rates (up to a logarithmic factor in the case of a zero of a polynomial order), over a wide range of Besov balls. Due to spatial irregularity, the estimators exhibit very different behavior and minimax convergence rates in comparison with the solution of spatially homogeneous ill-posed problem (see Remark 3). Specifically, even if $1/g$ is not integrable, one can recover f as well as in the case of an equispaced design (in terms of minimax convergence rates) when the function is homogeneous enough since the estimator is "borrowing strength" in the areas where f is adequately sampled. These features lead to a different structure of estimators of f described in Section 4. The complementary case when $1/g$ is integrable has been partially handled by Chesneau (2007a) who showed that the problem is well-posed (i.e., data loss does not affect convergence rates) when f is spatially homogeneous. The complementary case when $1/g$ is integrable which has been partially handled by Chesneau (2007a) is handled in Section 7. In depth discussion of the differences of the spatial features of spatially inhomogeneous ill-posed problem studied in this paper is presented in Section 8.

We limit our attention only to the case of an L^2-risk since the consideration of a wider class of risk functions will make the exposition of our work even longer; all results, however, obtained

can be extended to the case of L^u-risks, $1 \leq u < \infty$.

The rest of the paper is organized as follows. Section 2 discusses the formulation of the nonparametric regression estimation problem in the cases of moderate and severe data losses. In Section 3, we derive the asymptotical minimax lower bounds for the L^2-risk over a wide range of Besov balls. Section 4 talks about estimation strategies when $1/g$ is not integrable, in particular, about partitioning the unknown response function f and its estimator into the zero-affected and zero-free parts. Section 5 elaborates on the estimation of the zero-affected and the zero-free parts, and is followed by Section 6 which discusses the choice of adaptive resolution level and derives the asymptotical minimax upper bounds for the L^2-risk in the case when $1/g$ is not integrable. Section 7 studies complementary case when g has zeros but $1/g$ is still integrable. Section 8 concludes the paper with a discussion. Finally, Section 9 contains the proofs of the statements in the earlier sections.

2 Formulation of the problem

Consider the nonparametric regression model (1.1). Since the noise level is assumed to be known and finite, without loss of generality, we set $\sigma = 1$. Therefore, from now onwards, we work with observations y_i governed by equation (1.1) where $f \in L^2([0,1])$ is the unknown response function to be recovered, $x_i \in [0,1]$, $i = 1,2,\ldots,n$, are random design points with the underlying density function g, and ξ_i, $i = 1,2,\ldots,n$, are independent standard Gaussian random variables, independent of x_i, $i = 1,2,\ldots,n$. Furthermore, we assume that the design density g is known and has a finite number of well separated zeros on $[0,1]$. The last assumption is motivated by the following considerations. If g vanishes on an interval of non-asymptotic length, then consistent estimation of f is impossible. Also, the zeros of g have a concentration point only in the case when g is highly oscillatory, which is not a very likely scenario. Finally, the assumption that g has low values on a part of its domain but is still separated from zero is not an interesting case to consider, since the lower bound on g will appear in the constant of the well-known expressions for the minimax convergence rates (see, e.g., Tsybakov (2008)).

Note that the above assumptions are not restrictive. If the noise level σ is unknown, it can be easily estimated with parametric precision using observations in the region where g is separated from zero. The assumption that the design points x_i, $i = 1,2,\ldots,n$, are random is not confining either. In fact, with small modifications of the theory below, one can consider fixed points $0 \leq x_1 < x_2 < \cdots < x_n \leq 1$, generated by an increasing continuously differentiable function G such that $G(0) = 0$, $G(1) = 1$ and $G(x_i) = i/n$, $i = 1,2,\ldots,n$. Then, the function G plays the role of a "surrogate" distribution function with density function g; the design points x_i, $i = 1,2,\ldots,n$, can be then obtained as $x_i = G^{-1}(i/n)$, $i = 1,2,\ldots,n$.

Moreover, since the design density g is known and has only finite number of zeros which are well separated, one can partition interval $[0,1]$ into subintervals in such a manner that each subinterval contains only one zero of g. For this reason, without loss of generality, we assume that g has only one zero $x_0 \in (0,1)$ and the following conditions hold.

Assumption A. Let the design density function g be a continuous function on the interval $[0,1]$ with $g(x_0) = 0$, $x_0 \in (0,1)$. Then, there exists constants $\alpha \in \mathbb{R}$, $b \geq 0$ ($\alpha > 0$ if $b = 0$), $\beta > 0$ and $C_g > 0$ such that, for $x, x + x_0 \in [0,1]$,

$$\lim_{x \to 0} g(x_0 + x)|x|^{-\alpha} \exp(b|x|^{-\beta}) = C_g. \tag{2.1}$$

If $b = 0$, we shall say that x_0 is a zero of *polynomial order*. If $b > 0$, we shall say that x_0 is a zero of *exponential order*. Observe that (2.1) implies that there exist some absolute constants

$C_{g1} < C_g < C_{g2}$ such that for any x, with $x, x + x_0 \in [0, 1]$ and $x_0 \in (0, 1)$, one has

$$g(x_0 + x) \le C_{g2}|x|^\alpha \exp(-b|x|^{-\beta}), \quad g(x_0 + x) \ge C_{g1}|x|^\alpha \exp(-b|x|^{-\beta}). \tag{2.2}$$

Note that the two cases in Assumption A correspond to the situations of moderate ($b = 0$) and severe ($b > 0$) data losses, respectively. Chesneau (2007a) showed that in the case of a moderate loss with $0 < \alpha < 1$ (i.e., when $b = 0$ and $1/g$ is integrable) and a spatially homogeneous function, the unknown response function f can be estimated with the same asymptotical minimax convergence rates under the L^2-risk as in the case of $\alpha = 0$; hence, in this case, the nonparametric regression estimation problem turns out to be a *well-posed* problem.

We shall be, therefore, mainly interested only in the complementary situation when $1/g$ is not integrable: (i) moderate losses (i.e., $b = 0$) with $\alpha \ge 1$ and (ii) severe losses (i.e., $b > 0$) with $\alpha \in \mathbb{R}$ and $\beta > 0$. As we shall see below, usually in those cases, the asymptotically optimal estimation procedures yield estimators with lower convergence rates than in the case of equispaced observations, so that the corresponding nonparametric regression estimation problem under model (1.1) becomes *ill-posed* (see Remark 1), with the degree of ill-posedeness growing as $\alpha \ge 1$ increases when $b = 0$ or as $\beta > 0$ increases when $b > 0$.

In what follows, we use the symbol C for a generic positive constant, independent of the sample size n, which may take different values at different places.

Remark 1 (Risk functions and design). As indicated above, we shall measure the precision of any estimator $\hat{f}_n$ of f by its L^2-risk, i.e.,

$$\Delta(\hat{f}_n) = \mathbb{E}\|\hat{f}_n - f\|^2.$$

If the design points $x_i \in [0, 1]$, $i = 1, 2 \ldots, n$, are treated as fixed (i.e., non-random), then the above risk, evaluated at the equispaced design $\{i/n\}$, $i = 1, 2, \ldots, n$, corresponds to

$$\Delta^d(\hat{f}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\hat{f}_n(i/n) - f(i/n)]^2,$$

and leads to an *ill-posed* nonparametric regression estimation problem. However, it is instructive to note that if one measures the precision of an estimator $\hat{f}_n$ at the design points $x_i \in [0, 1]$, $i = 1, 2, \ldots, n$, only, by calculating

$$\Delta^d_{fixed}(\hat{f}_n, x_i) = \frac{1}{n}\sum_{l=1}^{n} \mathbb{E}[\hat{f}_n(x_i) - f(x_i)]^2,$$

as it was done in, e.g., Amato *et. al* (2006), then the problem ceases to be ill-posed. Moreover, in this case, no special treatment is necessary to account for the irregular design. To confirm that, note that equation (1.1) can be re-written as

$$y_i = F(i/n) + \xi_i, \quad i = 1, 2, \ldots, n, \tag{2.3}$$

where $F(x) = f(G^{-1}(x))$, $x \in [0, 1]$. Construct now an estimator $\hat{F}_n$ of F using, e.g., any of the standard wavelet thresholding techniques, and set $\hat{f}_n(x) = \hat{F}_n(G(x))$, $x \in [0, 1]$. Then,

$$\hat{F}_n(x) = \hat{f}_n(G^{-1}(x)), \quad x \in [0, 1],$$

and $\Delta^d_{fixed}(\hat{f}_n, x_i)$ takes the form

$$\Delta^d_{fixed}(\hat{f}_n, x_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\hat{F}_n(i/n) - F(i/n)]^2.$$

Therefore, if the observed data vector $y = (y_1, y_2, \dots, y_n)$ is treated as if the measurements were carried out at equispaced design points, then, by using, e.g., available wavelet denoising algorithms, the resulting estimator $\hat{F}_n$ of function F will be adaptive and it will lead to the smallest possible risk $\Delta^d_{fixed}(\hat{f}_n, x_i)$. This phenomenon was noticed earlier by Cai and Brown (1998), Sardy *et al.* (1999) and Brown *et al.* (2002).

Remark 2 (Local versus global convergence rates). The problem of estimating f globally is a much more difficult task than estimating f locally, say at a given point a. Indeed, if G is known, then $F(G(a)) = f(a)$ and, hence, one can estimate F at the point $G(a)$ instead of estimating f at the point a, where $F(x) = f(G^{-1}(x))$, $x \in [0,1]$, and F is uniformly sampled (see (2.3)). Hence, pointwise estimation can be reduced to a well-addressed pointwise regression estimation problem. If $g(a) \neq 0$, then the problem is well-posed and has been extensively studied before. If, instead, $a = x_0$ is a zero of g, then one can deduce minimax pointwise convergence rates directly from considerations of Remark 1 and straightforward calculus. Let, for simplicity, $x_0 = 0$ and $g(x) = \alpha x^{\alpha}$, so that $G(x) = x^{\alpha+1}$ and $G^{-1}(x) = x^{1/(\alpha+1)}$, $x \in [0,1]$. Let f satisfy a Hölder condition of order s at x_0, i.e., $|f(x) - f(x_0)| \leq C|x - x_0|^s$. Then, since $x_0 = 0$, $F(x) = f(G^{-1}(x))$, $x \in [0,1]$, satisfies a Hölder condition of order $s' = s/(\alpha+1)$ at 0, i.e.,

$$|F(x) - F(0)| \;=\; |f(G^{-1}(x)) - f(G^{-1}(x_0))| \leq C|G^{-1}(x) - G^{-1}(x_0)|^s = C|x - x_0|^{s/(\alpha+1)}.$$

Since $f(x_0) = F(0)$, one can set $\hat{f}(x_0) = \hat{F}(0)$ and obtain minimax pointwise convergence rates for $\hat{f}(x_0)$, on noting that

$$\mathbb{E}\|\hat{f}(x_0) - f(x_0)\|^2 = \mathbb{E}\|\hat{F}(0) - F(0)\|^2 = O\left(n^{-\frac{2s'}{2s'+1}}\right) = O\left(n^{-\frac{2s}{2s+\alpha+1}}\right),$$

which coincides with the minimax pointwise convergence rates obtained by Gaiffas (2005). The whole argument here rests on the fact that $f(x_0) = F(G(x_0))$, $x_0 \in [0,1]$, so one can estimate F instead of f at the respective point. This, however, cannot be accomplished when global estimation procedure is required since, in such a case, a Taylor expansion is needed, that can be applied only locally.

3 Minimax lower bounds for the L^2-risk over Besov balls

Before constructing an estimator of the unknown response function f under model (1.1), we first derive the asymptotical minimax lower bounds for the L^2-risk over a wide range of Besov balls.

Among the various characterizations of Besov spaces for $f \in L^p([0,1])$ in terms of wavelet bases, we recall that for an r-regular multiresolution analysis (see, e.g., Meyer, 1992, Chapter 2, pp 21–25), with $0 < s < r$, and for a Besov ball $B^s_{p,q}(A)$ defined as

$$B^s_{p,q}(A) = \{f \in L^p([0,1]) : \; f \in B^s_{p,q}, \; \|f\|_{B^s_{p,q}} \leq A\},$$

of radius $A > 0$ with $1 \leq p, q \leq \infty$, one has, with $s' = s + 1/2 - 1/p$,

$$B^s_{p,q}(A) = \left\{ f \in L^p([0,1]) : \left(\sum_{k=0}^{2^m-1} |a_{mk}|^p \right)^{1/p} + \left(\sum_{j=m}^{\infty} 2^{js'q} \left(\sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{q/p} \right)^{1/q} \leq A \right\}, \quad (3.1)$$

with respective sum(s) replaced by maximum if $p = \infty$ and/or $q = \infty$, where $s' = s + 1/2 - 1/p$ (see, e.g., Johnstone *et. al* (2004)). We study below the L^2-risk over Besov balls $B^s_{p,q}(A)$ defined as

$$R_n(B^s_{p,q}(A)) = \inf_{\tilde{f}_n} \sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\tilde{f}_n - f\|^2,$$

where $\|h\|$ is the L^2-norm of a function h defined on the unit interval, and the infimum is taken over all possible square-integrable estimators (i.e., measurable functions) $\tilde{f}_n$ of f based on observations y_i from model (1.1).

The following statement provides the asymptotical minimax lower bounds for the L^2-risk.

Theorem 1 *Let $1 \le p, q \le \infty$ and $s > \max(1/p, 1/2)$. Then, under Assumption A (with $\alpha > 0$ if $b = 0$, and $\alpha \in \mathbb{R}$ and $\beta > 0$ if $b > 0$), as $n \to \infty$,*

$$R_n(B^s_{p,q}(A)) \ge \begin{cases} C\, n^{-\frac{2s}{2s+1}} & \text{if} \quad b = 0,\ \alpha s < s', \\ C\, n^{-\frac{2s'}{2s'+\alpha}} & \text{if} \quad b = 0,\ \alpha s \ge s', \\ C\, (\ln n)^{-\frac{2s'}{\beta}} & \text{if} \quad b > 0. \end{cases} \tag{3.2}$$

Remark 3 (Global convergence rates). As we shall show below, the minimax global convergence rates in Theorem 1 are attainable for $b > 0$ and are attainable up to a logarithmic factor for $b = 0$. If $\alpha s = s'$, the minimax global convergence rates in the first and second parts of (3.2) coincide. Theorem 1 implies that, whenever $\alpha s \le s'$, the problem is not ill-posed, in a sense that the minimax global convergence rates are the same as in the case of an equispaced design. For $\alpha \ge 1$, this relation can take place only if $2 \le p \le \infty$, i.e., when the function is spatially homogeneous. In particular, $\alpha s \le s'$ holds true for any α such that $1 \le \alpha \le 1 + s^{-1}\,(1/2 - 1/p)$, i.e., when f is very homogeneous spatially (p is large, in particular, when $p > 2/[1 - (\alpha - 1)s]$ provided $1 < \alpha < 1 + 1/s$), so that even a relatively severe data loss does not lead to the reduction of minimax global convergence rates. If $0 < \alpha < 1$, then the problem is always well-posed whenever f is spatially homogeneous ($p \ge 2$) and also when f is spatially inhomogeneous ($1 \le p < 2$) and $0 < \alpha < 1 - (1/p - 1/2)/s$. Therefore, even if f is spatially inhomogeneous, the problem is well-posed whenever data loss is very limited ($0 < \alpha < 1 - (1/p - 1/2)/s$).

4 Estimation strategies when $1/g$ is not integrable

We consider a scaling function φ^* and a mother wavelet ψ^* that generate an orthonormal wavelet basis in $L^2(\mathbb{R})$, as those obtained from, e.g., an r-regular multiresolution analysis of $L^2(\mathbb{R})$, for some $r > 0$. We shall also assume that φ^* and ψ^* are both compactly supported, with integer bounds on their supports so that, for some $L_{\varphi^*},\ U_{\varphi^*},\ L_{\psi^*},\ U_{\psi^*} \in \mathbb{Z}$, with $L_{\varphi^*} < U_{\varphi^*}$, $L_{\psi^*} < U_{\psi^*}$,

$$\text{supp}(\varphi^*) = [L_{\varphi^*}, U_{\varphi^*}], \quad \text{supp}(\psi^*) = [L_{\psi^*}, U_{\psi^*}], \quad L_{\varphi^*} \le 0,\ U_{\varphi^*} \ge 0,\ U_{\varphi^*} - L_{\varphi^*} \ge 4.$$

(For instance, the Daubechies or Symmlets scaling functions φ^* and mother wavelets ψ^*, with filter number (number of vanishing moments) $N \ge 3$, satisfy (4.2) with $L_{\varphi^*} = 0$, $U_{\varphi^*} = 2N - 1$, $L_{\psi^*} = 1 - N$ and $U_{\psi^*} = N$, see, e.g., Mallat (1999), Section 7.2.)

We then obtain a periodized version of the wavelet basis on the unit interval, i.e., for $j \ge 0$ and $k = 0, 1, \ldots, 2^j - 1$, as

$$\varphi_{jk}(x) = \sum_{i \in \mathbb{Z}} 2^{j/2} \varphi^*(2^j(x + i) - k), \quad \psi_{jk}(x) = \sum_{i \in \mathbb{Z}} 2^{j/2} \psi^*(2^j(x + i) - k), \quad x \in [0, 1],$$

so that, for any $m \geq 0$, the set

$$\{\varphi_{mk},\ \psi_{jk}:\ j \geq m,\ k = 0, 1, \ldots, 2^j - 1\},$$

where

$$\varphi_{mk}(x) = 2^{m/2}\varphi(2^m x - k), \quad \psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), \quad x \in [0,1],$$

forms an orthonormal wavelet basis for $L^2([0,1])$ (see, e.g., Mallat (1999), Theorem 7.16). Hence, for any $m \geq 0$, any $f \in L^2([0,1])$, can be expanded as

$$f(x) = \sum_{k=0}^{2^m-1} a_{mk}\varphi_{mk}(x) + \sum_{j=m}^{\infty}\sum_{k=0}^{2^j-1} b_{jk}\psi_{jk}(x), \quad x \in [0,1], \tag{4.1}$$

where

$$a_{mk} = \int_0^1 f(x)\varphi_{mk}(x)\,dx, \quad k = 0, 1, \ldots, 2^m - 1,$$

$$b_{jk} = \int_0^1 f(x)\psi_{jk}(x)\,dx, \quad j \geq m, \quad k = 0, 1, \ldots, 2^j - 1.$$

Denote by $L_\varphi, U_\varphi, L_\psi$ and U_ψ the support bounds of the periodic scaling function φ and mother wavelet ψ. Note that the supports of φ^*_{mk} and φ_{mk} coincide if and only if $2^m > U_{\varphi^*} - L_{\varphi^*}$, and, similarly, the supports of ψ^*_{jk} and ψ_{jk} coincide if and only if $2^m > U_{\psi^*} - L_{\psi^*}$. Choose the lowest resolution level m_1 such that $2^{m_1} > \max\left(U_{\varphi^*} - L_{\varphi^*}, U_{\psi^*} - L_{\psi^*}\right)$, so that supports of periodic and non-periodic wavelets coincide. In this case, we obtain that

$$L_{\varphi^*} = L_\varphi,\ U_{\varphi^*} = U_\varphi,\ L_{\psi^*} = L_\psi,\ U_{\psi^*} = U_\psi,\ L_\varphi \leq 0,\ U_\varphi \geq 0,\ U_\varphi - L_\varphi \geq 4. \tag{4.2}$$

For any integer $l \geq 1$, denote $k_{0l} = 2^l x_0$. At each resolution level, we partition the set of all indices into the indices which are *zero–affected* and *zero–free*. In particular, let K^φ_{0m} and K^ψ_{0j} be the sets such that, for any integer $m \geq m_1$ and $j = m, m+1, \ldots$,

$$\begin{aligned} K^\varphi_{0m} &= \{k : 0 \leq k \leq 2^m - 1, L_\varphi - 1 < k_{0m} - k < U_\varphi + 1\}, \\ K^\psi_{0j} &= \left\{k : 0 \leq k \leq 2^j - 1, L_\psi - 1 < k_{0j} - k < U_\psi + 1\right\} \end{aligned}$$

and let

$$K^\varphi_{0mc} = \{k : 0 \leq k \leq 2^m - 1, k \notin K^\varphi_{0m}\}, \quad K^\psi_{0jc} = \left\{k : 0 \leq k \leq 2^m - 1, k \notin K^\psi_{0j}\right\}.$$

Simple calculations yield that $k \in K^\varphi_{0mc}$ and $k \in K^\psi_{0jc}$ imply that $x_0 \notin \text{supp}\ \varphi_{mk}$ and $x_0 \notin \text{supp}\ \psi_{jk}$, respectively, so that the sets K^φ_{0mc} and K^ψ_{0jc} are zero–free while the sets K^φ_{0m} and K^ψ_{0j} are zero–affected.

With the above notation it is easy to see that, for any $m \geq m_1$, f can be partitioned as the sum of zero–affected and zero–free parts, i.e.,

$$f(x) = f_{0,m}(x) + f_{c,m}(x), \quad x \in [0,1],$$

where

$$f_{0,m}(x) = \sum_{k \in K^\varphi_{0m}} a_{mk}\varphi_{mk}(x) + \sum_{j=m}^{\infty}\sum_{k \in K^\psi_{0j}} b_{jk}\psi_{jk}(x), \quad x \in [0,1], \tag{4.3}$$

$$f_{c,m}(x) = \sum_{k \in K^\varphi_{0mc}} a_{mk}\varphi_{mk}(x) + \sum_{j=m}^{\infty}\sum_{k \in K^\psi_{0jc}} b_{jk}\psi_{jk}(x), \quad x \in [0,1]. \tag{4.4}$$

We then construct estimators $\hat{f}_{0,m}$ and $\hat{f}_{c,m}$ of $f_{0,m}$ and $f_{c,m}$, respectively, and estimate f by

$$\hat{f}_m(x) = \hat{f}_{0,m}(x) + \hat{f}_{c,m}(x), \quad x \in [0,1]. \tag{4.5}$$

(We emphasize the unusual feature in the construction of $\hat{f}_m$: as we shall see below, $\hat{f}_{0,m}$ is a linear estimator while $\hat{f}_{c,m}$ is a non-linear estimator with the lowest resolution level m determined by the linear part $\hat{f}_{0,m}$.)

By observing that for any function $u(x)$ we have

$$\int_0^1 u(x)f(x)dx = \mathbb{E}\left(\frac{f(X)u(X)}{g(X)}\right),$$

and setting $u(x) = \varphi_{mk}(x)$ and $u(x) = \psi_{jk}(x)$, $x \in [0,1]$, in turn, similarly to Chesneau (2007a), we estimate a_{mk}, $k \in K^{\varphi}_{0mc}$, and b_{jk}, $k \in K^{\psi}_{0jc}$, respectively, by

$$\hat{a}_{mk} = \frac{1}{n}\sum_{i=1}^{n} \frac{\varphi_{mk}(x_i)y_i}{g(x_i)}, \quad k \in K^{\varphi}_{0mc}, \qquad \tilde{b}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \frac{\psi_{jk}(x_i)y_i}{g(x_i)}, \quad k \in K^{\psi}_{0jc}. \tag{4.6}$$

Note that since $1/g$ is not integrable, the estimators (4.6) would have infinite variances if $k \in K^{\varphi}_{0m}$ or $k \in K^{\psi}_{0j}$, so that one cannot construct an estimator of $f_{0,m}$ by direct estimation of wavelet coefficients. In this case, we shall use a linear estimator with the lowest resolution level m estimated from the data. In what follows, we shall consider the estimation of $f_{0,m}$ and $f_{c,m}$ separately.

5 Estimation of the zero-free and the zero–affected parts.

In order to estimate $f_{c,m}$, we construct a wavelet nonlinear thresholding estimator $\hat{f}_{c,m}$ as

$$\hat{f}_{c,m}(x) = \sum_{k \in K^{\varphi}_{0mc}} \hat{a}_{mk}\varphi_{mk}(x) + \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \hat{b}_{jk}\psi_{jk}(x), \quad m_1 \le m \le J-1, \; x \in [0,1], \tag{5.1}$$

where $\hat{a}_{mk}$ are given in (4.6), J is defined below in (5.3) while the coefficients $\hat{b}_{jk}$ are thresholded estimators of the wavelet coefficients b_{jk} defined as

$$\hat{b}_{jk} = \begin{cases} \tilde{b}_{jk}\, \mathbb{I}(\tilde{b}^2_{jk} > d^2 n^{-1} \ln n \, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}) & \text{if } \; b = 0, \\ \tilde{b}_{jk}\, \mathbb{I}(|k - k_{0j}| > 2^{j-m}) & \text{if } \; b > 0. \end{cases} \tag{5.2}$$

Here, $d > 0$ is a constant, $\tilde{b}_{jk}$ are defined by (4.6) and m is such that $m_1 \le m \le J-1$, where

$$2^{m_1} = \max\left(U_{\varphi^*} - L_{\varphi^*}, U_{\psi^*} - L_{\psi^*}\right) + 1, \qquad 2^J = \begin{cases} (n/\ln n)^{1/(\alpha+1)} & \text{if } \; b = 0, \\ (\ln n)^{2/\beta} & \text{if } \; b > 0. \end{cases} \tag{5.3}$$

Now, consider estimation of the zero-affected part. Since the estimators $\hat{a}_{mk}$ of a_{mk}, given in (4.6), have infinite variances when $k \in K^{\varphi}_{0m}$, we estimate those coefficients by solving a system of linear equations. Note that there is a finite known number of indices in K^{φ}_{0m}, at most, $w_\phi = U_\varphi - L_\varphi$ indices. For a given $m \in [m_1, J-1]$, denote

$$f_m(x) = \sum_{k=0}^{2^m-1} a_{mk}\varphi_{mk}(x), \quad \varepsilon_m(x) = \sum_{j=m}^{\infty} \sum_{k=0}^{2^j-1} b_{jk}\psi_{jk}(x), \quad x \in [0,1], \tag{5.4}$$

and observe that $f(x) = f_m(x) + \varepsilon_m(x)$, so that

$$\sum_{k \in K^{\varphi}_{0m}} a_{mk}\varphi_{mk}(x) = f_m(x) - \varepsilon_m(x) - \sum_{k \in K^{\varphi}_{0mc}} a_{mk}\varphi_{mk}(x), \quad x \in [0,1]. \tag{5.5}$$

Denote $\Omega_\delta = [L_\varphi + \delta_b, U_\varphi - \delta_b]$, and choose δ_b such that

$$\delta_b = \begin{cases} 0 < \delta_b < 1/2, \ \varphi(L_\varphi + \delta_b) \neq 0, \ \varphi(U_\varphi - \delta_b) \neq 0, & \text{if } \ b > 0, \\ 0, & \text{if } \ b = 0. \end{cases} \tag{5.6}$$

Introduce also a finite set of indices

$$K^*_{0m} = \{k : 0 \le k \le 2^m - 1, \ 2L_\varphi - U_\varphi \le k_{0m} - k < L_\varphi \ \text{ or } \ U_\varphi < k_{0m} - k \le 2U_\varphi - L_\varphi\}. \tag{5.7}$$

Now, multiply both sides of formula (5.5) by $g(x)\,\varphi_{ml}(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)$, $l \in K^{\varphi}_{0m}$, where $\mathbb{I}(x \in \Omega)$ is the indicator of set Ω, and integrate. As a result, obtain the following system of linear equations

$$\mathbf{A}^{(m)}\mathbf{u}^{(m)} = \mathbf{c}^{(m)} - \boldsymbol{\varepsilon}^{(m)} - \mathbf{B}^{(m)}\mathbf{v}^{(m)}. \tag{5.8}$$

Here, matrices $\mathbf{A}^{(m)}$ and $\mathbf{B}^{(m)}$ and vectors $\mathbf{c}^{(m)}$, $\boldsymbol{\varepsilon}^{(m)}$, $\mathbf{u}^{(m)}$ and $\mathbf{v}^{(m)}$ have, respectively, elements

$$A^{(m)}_{lk} = \int_0^1 \varphi_{mk}(x)\varphi_{ml}(x)g(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)dx, \quad k, l \in K^{\varphi}_{0m}, \tag{5.9}$$

$$B^{(m)}_{lk} = \int_0^1 \varphi_{mk}(x)\varphi_{ml}(x)g(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)dx, \quad l \in K^{\varphi}_{0m}, \ k \in K^*_{0m}, \tag{5.10}$$

$$c^{(m)}_l = \int_0^1 f(x)\varphi_{ml}(x)g(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)dx, \quad l \in K^{\varphi}_{0m}, \tag{5.11}$$

$$\varepsilon^{(m)}_l = \int_0^1 \varepsilon_m(x)\varphi_{ml}(x)g(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)dx, \quad l \in K^{\varphi}_{0m}, \tag{5.12}$$

$$u^{(m)}_k = a_{mk}, \quad k \in K^{\varphi}_{0m}, \qquad v^{(m)}_k = a_{mk}, \quad k \in K^*_{0m}. \tag{5.13}$$

Note that the matrices $\mathbf{A}^{(m)}$ and $\mathbf{B}^{(m)}$ are completely known, and also observe that $B^{(m)}_{lk} \neq 0$ only if $k \in K^*_{0m}$, since, for $k \notin K^*_{0m}$, one has $\varphi_{mk}(x)\varphi_{ml}(x) = 0$.

Since $K^*_{0m} \subset K^{\varphi}_{0mc}$, it follows from (5.13) that components $v^{(m)}_k$ of vector $\mathbf{v}^{(m)}$ can be estimated by

$$\hat{v}^{(m)}_k = \hat{a}_{mk}, \quad k \in K^{\varphi}_{0mc},$$

using (4.6). We also estimate $c^{(m)}_l$ by

$$\hat{c}^{(m)}_l = \frac{1}{n}\sum_{i=1}^{n} y_i\,\varphi_{ml}(x_i)\,\mathbb{I}(2^m x_i - l \in \Omega_\delta), \quad l \in K^{\varphi}_{0m}, \tag{5.14}$$

and ignore vector $\boldsymbol{\varepsilon}$ in (5.8), thus, replacing (5.8) by the following system of linear equations

$$\mathbf{A}^{(m)}\hat{\mathbf{u}}^{(m)} = \hat{\mathbf{c}}^{(m)} - \mathbf{B}^{(m)}\hat{\mathbf{v}}^{(m)}. \tag{5.15}$$

Since matrix $\mathbf{A}^{(m)}$ is a positive definite matrix of non-asymptotic size, $\det(\mathbf{A}^{(m)}) \neq 0$ and we obtain the solution

$$\hat{\mathbf{u}}^{(m)} = (\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)} - \mathbf{B}^{(m)}\hat{\mathbf{v}}^{(m)})$$

of the system of linear equations (5.15) and set $\hat{a}_{mk} = \hat{u}_k^{(m)}$, $k \in K_{0m}^{\varphi}$. Finally, for a given m, we set $\hat{a}_{mk} = \hat{u}_k^{(m)}$, $k \in K_{0m}^{\varphi}$, and estimate $f_{0,m}$ by the following wavelet linear estimator

$$\hat{f}_{0,m}(x) = \sum_{k \in K_{0m}^{\varphi}} \hat{a}_{mk} \varphi_{mk}(x), \quad x \in [0,1]. \tag{5.16}$$

The following statement provides the asymptotic upper bounds for the bias and the variance of the estimator $\hat{f}_{0,m}$ given in (5.16).

Lemma 1 *Denote $f_{0,m}(x) = \sum_{k \in K_{0m}^{\varphi}} a_{mk} \varphi_{mk}(x)$ and let $m = m(n)$ be a non-random quantity, $m(n) \to \infty$ as $n \to \infty$. Then, $\hat{f}_{0,m}(x)$ defined in (5.16) satisfies, as $n \to \infty$,*

$$\|\mathbb{E}\hat{f}_{0,m} - f_{0,m}\|^2 = O\left(2^{-2ms'}\right), \quad \mathbb{E}\|\hat{f}_{0,m} - \mathbb{E}\hat{f}_{0,m}\|^2 = O\left(n^{-1} 2^{m\alpha} \ \exp(b\, 2^{m\beta}[2^{\beta+1}+1])\right). \tag{5.17}$$

Moreover, if $b = 0$, then $\mathbb{E}\|\hat{f}_{0,m} - \mathbb{E}\hat{f}_{0,m}\|^4 = o(1)$.

Define m_0 to be such that

$$2^{m_0} = \begin{cases} n^{\frac{1}{2s'+\alpha}} & \text{if} \quad b = 0, \\ \left(b^{-1} 2^{-(\beta+2)} \ln n\right)^{\frac{1}{\beta}} & \text{if} \quad b > 0. \end{cases} \tag{5.18}$$

It follows from Lemma 1 that, if $m = m_0$, the error $\mathbb{E}\|\hat{f}_{0,m} - f_{0,m}\|^2$ of estimator $\hat{f}_{0,m}$ attains the lower bounds in Theorem 1. Since α, b and β in (5.18) are known, in the case of $b > 0$ the value of m_0 is known. Therefore, one can selects m_0 as the lowest resolution level in the estimator of the zero-free part (5.1). The following lemma demonstrates that the estimator $\hat{f}_{c,m}$, given in (5.1), indeed attains the minimax global convergence rates in this case.

Lemma 2 *Let $1 \le p, q \le \infty$ and $s \ge \max(1/2, 1/p)$. Then, under Assumption A, with $b > 0$ and $\alpha \in \mathbb{R}$, for the estimator (5.1) with $m = m_0$, as $n \to \infty$,*

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E}\|\hat{f}_{c,m_0} - f_{c,m_0}\|^2 \le C\, (\ln n)^{-\frac{2s'}{\beta}}. \tag{5.19}$$

Unfortunately, this idea cannot be implemented in the case of $b = 0$. Indeed, though α in (5.18) is known, the value of s' is unknown and, therefore, the estimator $\hat{f}_{0,m}$ with $m = m_0$ is not realizable if $b = 0$. In this case, we need to choose resolution level $\hat{m}$ which approximates m_0 in some sense and then estimate f by $\hat{f}(x) = \hat{f}_{0,\hat{m}}(x) + \hat{f}_{c,\hat{m}}(x)$. The choice of such resolution level is a rather difficult task. On the one hand, $\hat{m}$ should not be too small, otherwise, the linear portion of the estimator would have bias which is too large. On the other hand, since $\hat{f}_{0,m}$ is the linear estimator, in order to represent $f = f_{0,\hat{m}} + f_{c,\hat{m}}$ adequately, $\hat{m}$ has to be used as the lowest resolution level in $\hat{f}_{c,\hat{m}}(x)$. The following lemma provides upper bounds for the risk of the estimator (5.1) of the zero-free part $f_{c,m}$ when $b = 0$ and shows that the risk contains the component $n^{-1}\, 2^{m\alpha}$, so that in order to attain the minimax risk (3.2), one needs $\hat{m} \le m_0$ with high probability.

Lemma 3 *Let $1 \le p, q \le \infty$ and $s \ge \max(1/2, 1/p)$, and let Assumption A hold with $\alpha \ge 1$. Let $\hat{f}_{c,m}$ be given by (5.1) where the non-random quantity $m = m(n)$ is such that $m_1 \le m \le J-1$ with m_1 and J defined in (5.3). Let $\hat{b}_{jk}$ be given by (5.2), and $d > 4C_d$, where C_d is given by*

$$C_d = 8C_{\psi}\,(C_{g1})^{-1} \ \max\left(2,\ 2\|f\|_{\infty}^2,\ \|f\|_{\infty}\|\psi\|_{\infty}/3,\ \|\psi\|_{\infty}\right) \quad \textit{with} \quad C_{\psi} = [2\max(|L_{\psi}|, |U_{\psi}|)]^{\alpha}. \tag{5.20}$$

Then, for the estimator $\hat{f}_{c,m}$, defined in (5.1), as $n \to \infty$,

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^2 = \begin{cases} O\left(n^{-1}\, 2^{m\alpha}\, (\ln n)^{\mathbb{I}(\alpha=1)} + n^{-\frac{2s}{2s+1}}\, (\ln n)^{\mu_1}\right) & \text{if} \quad b=0,\ \alpha s < s', \\ O\left(n^{-1}\, 2^{m\alpha}\, (\ln n)^{\mathbb{I}(\alpha=1)} + n^{-\frac{2s'}{2s'+\alpha}}\, (\ln n)^{\mu_2}\right) & \text{if} \quad b=0,\ \alpha s \geq s'. \end{cases} \tag{5.21}$$

Here,

$$\mu_1 = \frac{2s(1+\mathbb{I}(\alpha=1))}{2s+1}, \quad \mu_2 = \frac{2s'(1+\mathbb{I}(\alpha=1))}{2s'+\alpha} + \mathbb{I}\left(\frac{s'}{s} = \alpha > 1\right).$$

Moreover, as $n \to \infty$,

$$\mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^4 = o\,(1)\,. \tag{5.22}$$

6 Adaptive estimators and the minimax upper bounds for the L^2-risk when $1/g$ is not integrable

In order to construct an adaptive wavelet estimator of f in the case of $b = 0$, we shall use the technique of optimal tuning parameter selection pioneered by Lepski (1990, 1991) and further exploited in Lepski and Spokoiny (1997) and Lepski *et al.* (1997). The idea behind this technique is to construct estimators for various values of the tuning parameter in question (m, in our case), and then choose an optimal value of the tuning parameter by regulating the differences between the estimators constructed with different values of the parameter.

In particular, if $b = 0$, for various values of m, we construct versions of the system of equations (5.15) where estimators $\hat{\mathbf{v}}^{(m)}$ are constructed as before, solve those systems and obtain estimators (5.16) with where $\hat{a}_{mk} = \hat{u}^{(m)}_k$, $k \in K^{\varphi}_{0m}$. Construct an estimator $\hat{f}_m$ of f using formula (4.5) where $\hat{f}_{0,m}$ and $\hat{f}_{c,m}$ are of the forms (5.16) and (5.1), respectively, and m is the lowest resolution level of $\hat{f}_{c,m}$. The choice of the optimal resolution level is driven by the zero-affected portion of f rather than the zero-free portion. For this reason, for any resolution level $m > 0$, we define a neighborhood Ξ_m of x_0 as

$$\Xi_m = \left\{x:\ 2^{-m}[\min(L_\varphi, L_\psi) - U_\varphi] < x - x_0 < 2^{-m}[\max(U_\varphi, U_\psi) - L_\varphi]\right\} \tag{6.1}$$

and observe that Ξ_m is designed so that $\mathrm{supp}(f_{0,m}) \subseteq \Xi_m$, $\mathrm{supp}(\hat{f}_{0,m}) \subseteq \Xi_m$ and $\Xi_j \subset \Xi_m$ if $j > m$.

For $b = 0$, choose $m = \hat{m}$ such that $m_1 \leq m \leq J-1$, where m_1 and J are defined in (5.3) and

$$\hat{m} = \min\ \left\{m:\quad \|(\hat{f}_m - \hat{f}_j)\mathbb{I}(\Xi_m)\|^2 \leq \lambda^2\, 2^{j\alpha}\, n^{-1} \ln n \ \text{ for all } \ j,\ m \leq j \leq J-1\right\}, \tag{6.2}$$

where $\lambda > 0$ is a constant to be defined below. For completeness, define

$$\hat{m} = m_0 \quad \text{if} \ \ b > 0, \tag{6.3}$$

where m_0 is defined in (5.18).

The construction of $\hat{m}$ for $b = 0$ is based on the following idea. Note that when $\hat{m} \leq m_0$, then for $m = \hat{m}$, one has

$$\mathbb{E}\|\hat{f}_m - f\|^2 \leq 2\left[\mathbb{E}\|\hat{f}_m - \hat{f}_{m_0}\|^2 + \mathbb{E}\|\hat{f}_{m_0} - f\|^2\right]. \tag{6.4}$$

The first component in (6.4) is small due to definition of the resolution level $\hat{m}$ while the second component is calculated at the optimal resolution level m_0 and, hence, tends to zero at the optimal convergence rate (up to a logarithmic factor). On the other hand, if $m = \hat{m} > m_0$, then there exists $j > m$ such that $\|(\hat{f}_m - \hat{f}_j)\mathbb{I}(\Xi_m)\|^2 > \lambda^2\, 2^{j\alpha}\, n^{-1} \ln n$. The following Lemma shows that, if λ is large enough, the probability of this event is infinitesimally small.

Lemma 4 *Let $b = 0$ and let m_0 and $\hat{m}$ be given by the expressions (5.18) and (6.2), respectively. Denote*

$$C_{\lambda 0} = 4\sqrt{2(U_\varphi - L_\varphi + 1)}, \quad C_{\lambda 1} = C_{\lambda 0}(\sqrt{2}\, C_{g2})^{-1}\, \|(\mathbf{A}^*)^{-1}\|, \quad C_{\lambda 2} = C_{\lambda 0}\, \|(\mathbf{A}^*)^{-1}\mathbf{B}^*\|. \tag{6.5}$$

Let C_d be given by (5.20) and

$$C_\lambda = \max\left(2C_u, C_\tau C_{\lambda 0}\right), \tag{6.6}$$

where

$$\begin{aligned} C_u &= \max\left(C_{\lambda 1}C_\kappa, C_{\lambda 2}C_\tau\right), & (6.7)\\ C_\tau &= 8C_\varphi\, (C_{g1})^{-1}\ \max\left(2, 2\|f\|_\infty^2, \|f\|_\infty\|\varphi\|_\infty/3, \|\varphi\|_\infty\right), & (6.8)\\ C_\kappa &= \min_{a>0}\max\left(16C_\varphi C_{g2}\|f\|_\infty, 16a, \frac{8\|f\|_\infty\|\varphi\|_\infty}{3}, 16C_\varphi C_{g2}, \frac{4C_\varphi C_{g2}\|\varphi\|_\infty}{a^2}, \frac{4\|\varphi\|_\infty^2}{3a}\right). & (6.9)\end{aligned}$$

Here $||h||_\infty$ is the uniform norm of a bounded function h defined on the unit interval, C_{g1} and C_{g2} are defined by (2.2) and $C_\varphi = [2\max(|L_\varphi|, |U_\varphi|)]^\alpha$. If $\lambda \geq \max\left(C_{\lambda 1}, C_{\lambda 2}\right)$, then, as $n \to \infty$,

$$\mathbb{P}\left(\hat{m} > m_0\right) = O\left(n^{-\frac{\lambda}{C_\lambda}}\right) + O\left(n^{\frac{1}{\alpha+1} - \frac{d}{2C_d}}\right). \tag{6.10}$$

Lemma 4 confirms that indeed $m = \hat{m}$ can be chosen as the lowest resolution level in the nonlinear portion of the estimator, so that we estimate f by

$$\hat{f}(x) = \hat{f}_{0,\hat{m}}(x) + \hat{f}_{c,\hat{m}}(x), \quad x \in [0,1], \tag{6.11}$$

where $\hat{f}_{0,m}(x)$ and $\hat{f}_{c,m}(x)$ are defined in (5.16) and (5.1), respectively. The following statement confirms that the wavelet nonlinear estimator $\hat{f}$ given by (6.11) indeed attains (up to a logarithmic factor) the asymptotic minimax lower bounds obtained in Theorem 1.

Theorem 2 *Let $1 \leq p, q \leq \infty$ and $s \geq \max(1/2, 1/p)$ and Assumption A hold with $\alpha \geq 1$ if $b = 0$ and $\alpha \in \mathbb{R}$ if $b > 0$. Let $\hat{f}$ be the wavelet estimator defined by (6.11) with $\lambda > \max\left(2C_\lambda, C_{\lambda 1}, C_{\lambda 2}\right)$ in (6.2) and $d > 2(\alpha+1)^{-1}(2\alpha+3)C_d$ where C_λ is defined in (6.6), $C_{\lambda 1}, C_{\lambda 2}$ are defined in (6.5), and C_d is defined in (5.20). Then, as $n \to \infty$,*

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f} - f\|^2 \leq \begin{cases} C\, n^{-\frac{2s}{2s+1}}\, (\ln n)^{\frac{2s(1+\mathbb{I}(\alpha=1))}{2s+1}} & \text{if} \quad b = 0,\ \alpha s < s', \\ C\, n^{-\frac{2s'}{2s'+\alpha}}\, (\ln n)^{\frac{2s'(1+\mathbb{I}(\alpha=1))}{2s'+\alpha} + \mathbb{I}\left(\frac{s'}{s} = \alpha > 1\right)} & \text{if} \quad b = 0,\ \alpha s \geq s', \\ C\, (\ln n)^{-\frac{2s'}{\beta}} & \text{if} \quad b > 0. \end{cases}$$

Remark 4 (**Adaptivity**) Theorems 1 and 2 demonstrate that, for severe data losses ($b > 0$), the adaptive wavelet nonlinear estimator $\hat{f}$ given by (6.11) attains the asymptotically optimal (in the minimax sense) global convergence rates. For moderate data losses ($b = 0$ with $\alpha \geq 1$), however, the adaptive wavelet nonlinear estimator $\hat{f}$ given by (6.11) is asymptotically near-optimal up to a logarithmic factor. Moreover, if p is large and $\alpha > 1$ is relatively small ($1 < \alpha < (1/2 - 1/p)/s$), data loss does not affect the minimax global convergence rates and they coincide with the minimax global convergence rates obtained in the absence of data losses.

Remark 5 (Relation to work of Gaïffas). Estimation of the "zero-affected" part of f in the paper is somewhat similar to the procedure of Gaïffas with the difference that he used local polynomials while we are using wavelets. However, the significant difference is that we use this estimator only for the zero-affected part and not for the whole function f.

Another difference between our and Gaïffas' studies is that, first, we are able to formulate the convergence rates explicitly, in a simple meaningful way, and, due to the fact that we are using thresholding of wavelet coefficients rather than solution of the system of equations as in Gaïffas, our estimator can adapt to the case when the estimated function is spatially inhomogeneous. Moreover, we should point out that our optimal convergence rates are global and over a wide range of Besov balls compare to Gaiffas that are local or uniform and only for Hölder spaces. In particular, Gaïffas (2005, 2007) deals only with estimation of f at x_0, zero of the design density. The rates of convergence of the estimator $\hat{f}(x_0)$ can be expressed explicitly via α and parameters of Hölder ball where f belongs. However, as we pointed out in Remark 2 , this problem is much easier than the global estimation and can be solved by straightforward calculus.

Gaïffas (2006, 2009) also studied convergence rates of his estimator using uniform norm. In this case, for example, the convergence rates are formulated in terms of a solution of a nonlinear equation. The are no explicit expressions for the rates in a general situation. For instance, the only example, which appears in Gaïffas (2009), is produced for the simplest situation when $\sigma = 1$, f belongs to a Hölder class with parameters $s = L = 1$ and the design density is of the form $g(x) = 4|x - 1/2|$, i.e. $\alpha = 1$. In this case, the convergence rates are given by

$$r_n(x) = (\log n/n)^{\alpha_n(x)},$$

where

$$\alpha_n(x) = \begin{cases} \frac{1}{3}\left(1 - \frac{1-2x)}{\log(\log n/n)}\right), & x \in [0, 0.5 - (\log n/2n)^{1/4}), \\ \frac{\left([(x-0.5)^4 + 4\log n/n]^{1/2} - (x-0.5)^2\right) - \log 2}{2\log(\log n/n)}, & x \in [0.5 - (\log n/2n)^{1/4}, 0.5 - (\log n/2n)^{1/4}], \\ \frac{1}{3}\left(1 - \frac{2x-1)}{\log(\log n/n)}\right), & x \in (0.5 + (\log n/2n)^{1/4}, 1] \end{cases}$$

In the cases, when $\alpha > 1$ or it is not an integer, a solution of the equation which produces the convergence rates, as well as derivation of the explicit expression for the rates, require very nontrivial investigation.

7 Estimation and the minimax upper bounds for the L^2-risk when $1/g$ is integrable

The case when $1/g$ is integrable, i.e. when g has a zero of a polynomial order α, $0 < \alpha < 1$, has been considered by Chesneau (2007a) who demonstrated that the problem is well-posed when f is spatially homogeneous, i.e., when $p \geq 2$. However, the lower bounds in Theorem 1 show that the problem becomes ill-posed when $\alpha s > s'$, i.e., when $1 \leq p < [s(1-\alpha) + 1/2]^{-1}$. Hence, by considering only spatially homogeneous regression functions ($p \geq 2$), Chesneau (2007a) missed the "elbow rate" when f is spatially inhomogeneous and the fact that the problem becomes ill-posed in this case. However, since the estimators of the wavelet coefficients (4.6) have finite variances for $0 < \alpha < 1$, one can construct an estimator similar to Chesneau (2007a) by simply thresholding wavelet coefficients. In particular, set

$$\hat{f}(x) = \sum_{k=0}^{2^{m_1}-1} \hat{a}_{m_1 k}\varphi_{m_1 k} + \sum_{j=m_1}^{J-1}\sum_{k=0}^{2^j-1} \hat{b}_{jk}\psi_{jk}(x), \tag{7.1}$$

where $\hat{a}_{mk}$ and $\hat{b}_{jk}$ are defined in (5.2), and m_1 and J are defined in (5.3) with $b = 0$. The following statement confirms that estimator (7.1) attains (up to a logarithmic factor) minimax lower bounds obtained in Theorem 1.

Theorem 3 *Let $1 \leq p < \infty, 1 \leq q \leq \infty$, $s \geq \max(1/2, 1/p)$, and let d in (5.2) satisfy inequality*

$$d > \frac{2C_d(3\alpha + 5)}{(1-\alpha)(1+\alpha)}, \tag{7.2}$$

where C_d is given by (5.20). Then, under Assumption A, with $0 < \alpha < 1$ and $b = 0$, as $n \to \infty$,

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f}_c - f\|^2 = \begin{cases} O\left(n^{-\frac{2s}{2s+1}} (\ln n)^{\frac{2s}{2s+1}}\right) & \text{if} \quad \alpha s < s', \\ O\left(n^{-\frac{2s'}{2s'+\alpha}} (\ln n)^{\frac{2s'}{2s'+\alpha} + \mathbb{I}(\alpha s \geq s')}\right) & \text{if} \quad \alpha s \geq s'. \end{cases} \tag{7.3}$$

Theorem 3 shows that for $0 < \alpha < 1$, the problem is regular a long as $p > [s(1-\alpha)+1/2]^{-1}$ and it becomes ill-posed when $p < [s(1-\alpha)+1/2]^{-1}$. Therefore, even when data loss is very moderate ($0 < \alpha < 1$), the problem becomes ill-posed whenever f is rather spatially non-homogeneous ($p < [s(1-\alpha)+1/2]^{-1}$).

8 Discussion

We considered the nonparametric regression estimation problem of recovering an unknown response function f on the unit interval $[0, 1]$ on the basis of incomplete data when the design density g is known and has a zero $x_0 \in (0, 1)$ of a polynomial or an exponential order. We investigated the global estimation (in the minimax sense) of f which is a much harder problem than pointwise estimation studied by Gaiffas (2005, 2007) since the problem cannot be reduced to the estimation of a related regularly-sampled function (see Remarks 1 and 2).

The problem of global nonparametric estimation of a regression function is ill-posed and, moreover, it is spatially inhomogeneous. For this reason, the resulting estimators demonstrate completely different patterns of behavior in comparison with spatially homogeneous ill-posed problems like, e.g., deconvolution.

We studied various regimes of data loss from relatively minor (when the design density g has a zero of polynomial order $\alpha \in (0, 1)$ and, therefore, $1/g$ is integrable) to moderate (when the design density g has a zero of polynomial order $\alpha \geq 1$ and, hence, $1/g$ is not integrable) and severe (when the design density g has a zero of exponential order, so that $1/g$ is not integrable).

Global convergence rates in the case of minor data losses ($0 < \alpha < 1$) were studied by Chesneau (2007a) who showed that the problem is well-posed (the minimax global convergence rates are the same as in the absence of data loss) whenever the regression function f is spatially homogeneous. As our study shows, the problem remains well posed even if f is spatially inhomogeneous as long as the data loss is very minor ($0 < \alpha < 1 - (1/p - 1/2)/s$) or the function is relatively smooth ($p > (1/2 - s(\alpha - 1))^{-1}$). When $\alpha \geq 1 - (1/p - 1/2)/s$ ($p \leq (1/2 - s(\alpha - 1))^{-1}$), the problem becomes ill-posed.

Now, consider the situation when data loss is moderate ($b = 0$ and the zero of g is of a polynomial order $\alpha \geq 1$). The problem is ill-posed if $\alpha \geq (1/2 - 1/p)/s$, i.e., it is always ill-posed when f is spatially inhomogeneous ($1 \leq p < 2$). However, as Remark 3 points out, when f is very spatially homogeneous (p is rather large) and data loss is relatively moderate ($1 < \alpha < (1/2 - 1/p)/s$), the problem of estimation of f ceases to be ill-posed and exhibits minimax global convergence rates observed when g is bounded from below. Thus, in the case when f is very spatially homogeneous, the estimator of f is "borrowing strength" in the areas where f is adequately

sampled and exhibits minimax global convergence rates common for regularly spaced regression estimation problems. This is very dissimilar to spati ally homogeneous ill-posed problems (e.g., deconvolution) where there is a change point in the minimax global convergence rates (the, so-called, elbow effect) when the estimated function $f \in B^s_{p,q}$ is spatially inhomogeneous ($1 \le p < 2$) and they are independent of p when it is spatially homogeneous ($2 \le p \le \infty$). On the contrary, in the case of spatially inhomogeneous ill-posed problems, like the one considered herein, the minimax global convergence rates depend on p even when the function is spatially homogeneous ($2 < p \le \infty$) as long as $\alpha \ge (1/2 - 1/p)/s$. Thus, the elbow effect occurs when $p > 2$, in particular, when $p > 2/(1 - (\alpha - 1)s)$ provided $1 < \alpha < 1 + 1/s$.

In the case when data loss is severe ($b > 0$ and the zero of g is of an exponential order $\beta > 0$), the minimax global convergence rates grow with p, i.e., the more spatially homogeneous f is, the better it can be estimated. This is unlike spatially inhomogeneous ill-posed problems, like a deconvolution problem, where the minimax global convergence rates improve when p is growing when $1 < p < 2$ and are independent of p when f is spatially homogeneous ($2 \le p \le \infty$).

The unusual behavior of the minimax global convergence rates in the case of the spatially inhomogeneous ill-posed problem considered above calls for different estimation strategies. In particular, whenever data loss is moderate or severe, we partition f into zero-affected and zero-free parts. First, we construct an adaptive linear wavelet estimator of the zero-affected part where the lowest resolution level $m = \hat{m}$ is independent of the unknown parameters of the Besov balls and, therefore, known when $b > 0$, and is chosen using Lepskii's method when $b = 0$. After that, we construct a nonlinear wavelet estimator of the zero-free part of f starting from the lowest resolution level $m = \hat{m}$. Note that nonlinear estimator is required even if g has a zero of exponential order ($b > 0$). This is very different from the case of spatially homogeneous ill-posed problems (e.g., deconvolution), where in the case of exponentially growing eigenvalues, a linear estimator usually attains optimal (in the minima sense) global convergence rates (see Pensky and Sapatinas (2009, 2010).)

We should mention that there is a significant difference between minimax local and minimax global convergence rates. Note that minimax local convergence rates at zero of g are always affected by loss of data, even for moderate data losses. The minimax global convergence rates, however, are not affected when data loss is limited and the regression function is very spatially homogeneous ($1 < \alpha < 1 + 1/s$ and $p > 2/(1 - (\alpha - 1)s)$. Finally, we point out that some of the logarithmic factors which appear in Theorems 1,2 and 3 could be removed by using block thresholding rather than term-by-term thresholding of wavelet coefficients.

Furthermore, due to its construction, the suggested wavelet estimator is not easily computable, so it is of limited practical use. Therefore, it is desirable to construct an alternative, more computational feasible, adaptive estimator which attains the asymptotically minimax global convergence rates, that was the aim of this work. This is the project for future work that we hope to address elsewhere.

Acknowledgements

Marianna Pensky was supported in part by National Science Foundation (NSF), grant DMS-1106564. The authors want to thank Yuri Golubev for valuable discussion of results and methodologies used in this paper.

9 Proofs

9.1 Lower bounds

Proof of Theorem 1. On noting that the asymptotical lower bound in Theorem 3.1 of Chesneau (2007a) is also true when $b = 0$ and $1/g$ is not integrable (i.e., $\alpha \geq 1$), the asymptotical lower bound in the first part of (3.2) can be obtain by the arguments of Chesneau (2007a) and, hence, we need to prove only the asymptotical lower bounds in the second and third parts of (3.2). For this purpose, we consider functions f_{jk} be of the form $f_{jk} = \gamma_j \psi_{jk}$ and let $f_0 \equiv 0$. Note that by (3.1), in order $f_{jk} \in B^s_{p,q}(A)$, we need $\gamma_j \leq A 2^{-js'}$. Set $\gamma_j = c 2^{-js'}$, where c is a positive constant such that $c < A$, and apply the following classical lemma on lower bounds:

Lemma 5 (Härdle, Kerkyacharian, Picard & Tsybakov (1998), Lemma 10.1). *Let V be a functional space, and let $d(\cdot,\cdot)$ be a distance on V. For $f, g \in V$, denote by $\Lambda_n(f,g)$ the likelihood ratio $\Lambda_n(f,g) = d\mathbb{P}_{X_n^{(f)}}/d\mathbb{P}_{X_n^{(g)}}$, where $d\mathbb{P}_{X_n^{(h)}}$ is the probability distribution of the process X_n when h is true. Let V contains the functions $f_0, f_1, \ldots, f_\aleph$ such that*

(a) $d(f_k, f_{k'}) \geq \delta > 0$ *for* $k = 0, 1, \ldots, \aleph$, $k \neq k'$,

(b) $\aleph \geq \exp(\lambda_n)$ *for some* $\lambda_n > 0$,

(c) $\ln \Lambda_n(f_0, f_k) = u_{nk} - v_{nk}$, *where v_{nk} are constants and u_{nk} is a random variable such that there exists $\pi_0 > 0$ with $\mathbb{P}_{f_k}(u_{nk} > 0) \geq \pi_0$,*

(d) $\sup_k v_{nk} \leq \lambda_n$.

Then, for an arbitrary estimator $\hat{f}$,

$$\sup_{f \in V} \mathbb{P}_{X_n^{(f)}}\big(d(\hat{f}, f) \geq \delta/2\big) \geq \pi_0/2.$$

Let now $V = \{f_{jk} : |k - k_{0j}| \leq K/2\}$, where $K > 2$ is a fixed positive constant, so that $\aleph = K$. Choose $d(f,g) = \|f - g\|$, where, as before, $\|\cdot\|$ denotes the L^2-norm on the interval $[0,1]$. Then, $d(f_{jk}, f_0) = \gamma_j = \delta$. Let $v_{nk} = \lambda_n = \ln K$ and $u_{nk} = \ln \Lambda_n(f_0, f_{jk}) + \ln K$. Now, in order to apply Lemma 5, we need to show that for some $\pi_0 > 0$, uniformly for all f_{jk}, we have

$$\mathbb{P}_{f_{jk}}(u_{nk} > 0) = \mathbb{P}_{f_{jk}}\left(\ln \Lambda_n(f_0, f_{jk}) > -\ln K\right) \geq \pi_0 > 0.$$

Since, by Chebychev's inequality,

$$\mathbb{P}_{f_{jk}}\left(\ln \Lambda_n(f_0, f_{jk}) > -\ln K\right) \geq 1 - \frac{\mathbb{E}_{f_{jk}}\big|\ln \Lambda_n(f_0, f_{jk})\big|}{\ln K},$$

we need to find a uniform upper bound for $\mathbb{E}_{f_{jk}}|\ln \Lambda_n(f_0, f_{jk})|$.

Note that

$$-2\ln \Lambda_n(f_0, f_{jk}) = \sum_{i=1}^n \gamma_j^2 \psi_{jk}^2(x_i) + 2\sum_{i=1}^n \gamma_j \psi_{jk}(x_i)\xi_i$$

where ξ_i, $i = 1, 2, \ldots, n$, are independent standard Gaussian random variables. Thus,

$$\mathbb{E}| - 2\ln \Lambda_n(f_0, f_{jk})| \leq A_n + 2B_n,$$

where

$$A_n = \mathbb{E}|\sum_{i=1}^n \gamma_j^2 \psi_{jk}^2(x_i)| = n\gamma_j^2 \int_0^1 \psi_{jk}^2(x) g(x) dx, \quad B_n = \mathbb{E}|\sum_{i=1}^n \gamma_j \psi_{jk}(x_i)\xi_i|.$$

Note that by Jensen's inequality,

$$
\begin{aligned}
B_n &= \mathbb{E}\left\{\mathbb{E}\left[\Big|\sum_{i=1}^{n}\gamma_j\psi_{jk}(x_i)\xi_i\Big|\;\Big|x_1,x_2,\ldots,x_n\right]\right\} \\
&\leq \mathbb{E}\left\{\mathbb{E}\left[\left(\sum_{i=1}^{n}\gamma_j\psi_{jk}(x_i)\xi_i\right)^2\Big|x_1,x_2,\ldots,x_n\right]\right\}^{1/2} = \sqrt{A_n},
\end{aligned}
$$

so that one needs uniform upper bounds for A_n only.

If j is large enough, A_n can be presented as

$$
A_n = n\gamma_j^2\int_{L_\psi}^{U_\psi}\psi^2(z)g(x_0+2^{-j}(z+k-k_{0j}))dz,
$$

where $k_{0j}=2^jx_0$. Observe that condition (2.1) implies that one has

$$
g(x_0+x)\leq C|x|^\alpha\exp(-b|x|^{-\beta}).
$$

Let $M_\psi=\max(|L_\psi|,|U_\psi|)$. Then, for a finite value of K, one has

$$
A_n \leq Cn\gamma_j^2 2^{-j\alpha}\,(M_\psi^\alpha+K^\alpha)\exp\left(-b2^{j\beta}(M_\psi+K)^{-\beta}\right).
$$

Now, recall that $\gamma_j=c2^{-js'}$ and choose the smallest possible value of j such that A_n are uniformly bounded. Simple calculation yield that

$$
A_n = O\left(2^{-j(2s'+\alpha)}\exp\left(-b2^{j\beta}[M_\psi+K]^{-\beta}\right)\right),
$$

so that $2^j=O\left(n^{1/(2s'+\alpha)}\right)$ if $b=0$ and $2^j=O\left((\ln n)^{1/\beta}\right)$ if $b>0$.

Now, applying Lemma 5 and Chebyshev inequality, we finally obtain

$$
\inf_{\tilde{f}_n}\sup_{f\in B^s_{p,q}(A)}\mathbb{E}\|\tilde{f}_n-f\|^2 \geq \inf_{\tilde{f}_n}\sup_{f\in V}(\gamma_j^2/4)\,\mathbb{P}(\|\tilde{f}_n-f\|>\gamma_j/2)\geq \pi_0\gamma_j^2/8,
$$

which, on noting that

$$
\frac{2s'}{2s'+\alpha}<\frac{2s}{2s+1}\quad\text{if and only if}\quad s'<\alpha s, \tag{9.1}
$$

completes the proof of the theorem.

9.2 Properties of estimators of wavelet coefficients

Consider the quantity

$$
J_{mkl}=\int 2^m|\varphi(2^mx-k)\varphi(2^mx-l)|\,g^{-1}(x)\,dx. \tag{9.2}
$$

Lemma 6 *Let $m=m(n)$ be a nonrandom value and let $\hat{a}_{mk}$ be defined by (4.6). Then, for $k,l\in K^\varphi_{0mc}$, as $n\to\infty$,*

$$
|Cov(\hat{a}_{mk},\hat{a}_{ml})| = O\left(n^{-1}\,(J_{mkl}+1)\right), \tag{9.3}
$$

where

$$
J_{mkl}=O\left(n^{-1}\,2^{m\alpha}|k-k_{0m}|^{-\alpha}\exp(b2^{m\beta}|k-k_{0m}|^{-\beta})\right)\quad \textit{if}\;\;|k-l|\leq U_\varphi-L_\varphi, \tag{9.4}
$$

and $J_{mkl} = 0$ otherwise. Moreover, if $b = 0$, then, as $n \to \infty$,

$$\begin{aligned} Var(\hat{a}_{mk}) &= O\left(n^{-1}\, 2^{m\alpha} |k - k_{0m}|^{-\alpha}\right), \\ \mathbb{E}(\hat{a}_{mk} - a_{mk})^4 &= O\left(n^{-3}\, 2^{m(3\alpha+1)} |k - k_{0m}|^{-3\alpha}\right) + O\left(n^{-2}\, 2^{2m\alpha} |k - k_{0m}|^{-2\alpha}\right). \end{aligned} \tag{9.5}$$

Similarly, if $k, l \in K^{\psi}_{0jc}$ and $b = 0$, then $\tilde{b}_{jk}$, defined in (4.6), satisfy, as $n \to \infty$,

$$Var(\tilde{b}_{jk}) = O\left(n^{-1}\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}\right), \tag{9.6}$$

$$\mathbb{E}(\tilde{b}_{jk} - b_{jk})^4 = O\left(n^{-3}\, 2^{j(3\alpha+1)} |k - k_{0j}|^{-3\alpha}\right) + O\left(n^{-2}\, 2^{2j\alpha} |k - k_{0j}|^{-2\alpha}\right), \tag{9.7}$$

$$\begin{aligned} \mathbb{E}(\tilde{b}_{jk} - b_{jk})^6 &= O\left(n^{-5}\, 2^{j(5\alpha+2)} |k - k_{0j}|^{-5\alpha}\right) + O\left(n^{-4}\, 2^{j(4\alpha+1)} |k - k_{0j}|^{-4\alpha}\right) \\ &\quad + O\left(n^{-3}\, 2^{3j\alpha} |k - k_{0j}|^{-3\alpha}\right). \end{aligned} \tag{9.8}$$

Proof of Lemma 6. Let us first prove formula (9.4). Changing variables $z = 2^m(x - x_0)$ in the last integral, and using inequality (2.2), derive that

$$J_{mkl} \le \frac{2^{m\alpha}}{C_{g1}} \int_{L_\varphi}^{U_\varphi} \frac{|\varphi(z)||\varphi(z + k - l)|\; dz}{|z + k - k_{0m}|^{\alpha} \exp\left(-b 2^{m\beta} |z + k - k_{0m}|^{-\beta}\right)}.$$

It is easy to note that $J_{mkl} = 0$ if $|k - l| > U_\varphi - L_\varphi$. Also, $k \in K^{\psi}_{0jc}$ implies that $k_{0m} - k \le L_\varphi - 1$ or $k_{0m} - k \ge U_\varphi + 1$, so that one has $|z + k - k_{0m}| \ge 1$ and, hence, $|z + k - k_{0m}| \propto |k - k_{0m}|$ which proves (9.4). Now, by direct calculations we obtain that

$$\mathrm{Cov}(\hat{a}_{mk}, \hat{a}_{ml}) = n^{-1} \left\{ \int [\sigma^2 + f^2(x)]\; \varphi_{mk}(x) \varphi_{ml}(x) g^{-1}(x) dx - a_{mk} a_{ml} \right\},$$

so that (9.3) is valid.

Since the proofs for the scaling and the wavelet coefficients in Lemma 6 are similar, we shall prove only formulae (9.6) – (9.8). Observe that, due to (2.2) and the fact that $k \in K^{\psi}_{0jc}$ implies $k_{0j} - k \le L_\psi - 1$ or $k_{0j} - k \ge U_\psi + 1$, by considerations similar to the ones provided above, for integers $r_1, r_2 > 0$, one has

$$\int (g(x))^{-r_2}\, (\psi_{jk}(x))^{2r_1}\, dx \le C\, 2^{j(r_1 - 1)}\; 2^{j r_2 \alpha}\; |k - k_{0j}|^{-r_2 \alpha}. \tag{9.9}$$

Now, to complete the proof of (9.6)–(9.8), as $n \to \infty$, apply (9.9) to the following formulae

$$\begin{aligned} \mathrm{Var}(\tilde{b}_{jk}) &= O\left(n^{-1} \int g^{-1}(x) \psi^2_{jk}(x) dx\right), \\ \mathbb{E}(\tilde{b}_{jk} - b_{jk})^4 &= O\left(n^{-3} \int g^{-3}(x) \psi^4_{jk}(x) dx + n^{-2} \left[\int g^{-1}(x) \psi^2_{jk}(x) dx\right]^2\right), \\ \mathbb{E}(\tilde{b}_{jk} - b_{jk})^6 &= O\left(n^{-5} \int g^{-5}(x) \psi^6_{jk}(x) dx + n^{-3} \left[\int g^{-1}(x) \psi^2_{jk}(x) dx\right]^3 \right. \\ &\quad \left. +\; n^{-4} \int g^{-3}(x) \psi^4_{jk}(x) dx \int g^{-1}(x) \psi^2_{jk}(x) dx \right). \end{aligned}$$

9.3 Proofs of the supplementary statements used in the proof of Lemma 1.

Lemma 7 *Let $\delta_0 = 0.5\ \ 3^{\beta+1}\ [2\ 3^{\beta+1} + (2M)^{\beta+1}]^{-1}$. Let $0 < \delta < \delta_0$ and $a, b \in (2-\delta, M]$ for some $M > 0$. Let $c > 0$ be such that $c < \min(a,b) + \delta$ and $c < \max(a,b) - (1-2\delta)$. Then,*

$$a^{-\beta} + b^{-\beta} - 2c^{-\beta} \le -0.5\,\beta M^{-(\beta+1)}.$$

Proof of Lemma 7. To show that the lemma is true, note that $\delta < \delta_0$ implies $\delta_0 < 1/2$. Let, without loss of generality, $a \le b$. Then, $c \le a + \delta$, $c \le b - (1-2\delta)$ and

$$\begin{aligned} a^{-\beta} - c^{-\beta} &\le a^{-(\beta+1)}\beta\delta \le (2-\delta)^{-(\beta+1)}\beta\delta \\ b^{-\beta} - c^{-\beta} &\le b^{-\beta} - (b-1+2\delta)^{-\beta} \le -\beta M^{-(\beta+1)}\,(1-2\delta). \end{aligned}$$

Therefore, taking into account that $2 - \delta \ge 3/2$ and $\delta < \delta_0$, we obtain

$$\begin{aligned} & a^{-\beta} + b^{-\beta} - 2c^{-\beta} \le (2-\delta)^{-(\beta+1)}\beta\delta - \beta M^{-(\beta+1)}\,(1-2\delta) \\ & \quad \le -\beta M^{-(\beta+1)}(3/2)^{-(\beta+1)}[(3/2)^{\beta+1} - \delta(M^{\beta+1} + 2(3/2)^{\beta+1})] \le -0.5 M^{-(\beta+1)}, \end{aligned}$$

which proves the lemma.

Lemma 8 *Let $A_{kl}^{(m)}$, $B_{kl}^{(m)}$, $c_l^{(m)}$ and $\hat{c}_l^{(m)}$ be given by (5.9), (5.10), (5.11) and (5.14), respectively. Then, $Var(\hat{c}_k^{(m)}) = O\left(n^{-1}\, A_{kk}^{(m)}\right)$, and for some absolute constants C_1 and C_2 one has*

$$C_1 2^{-m\alpha}\exp(-b2^{\beta(m+1)}) \le A_{kk}^{(m)} \le C_2 2^{-m\alpha}\exp(-b\, M_\varphi^{-\beta}\ 2^{m\beta}), \tag{9.10}$$

where $M_\varphi = U_\varphi - L_\varphi + \max(|U_\varphi|, |L_\varphi|)$. Moreover, if $b > 0$ and $0 < \delta_b < \delta_0$ for $\delta_0 = 0.5\ \ 3^{\beta+1}\ [2\ 3^{\beta+1} + (U_\varphi + L_\varphi)^{\beta+1}]^{-1}$, then

$$\frac{|A_{kl}^{(m)}|}{\sqrt{A_{kk}^{(m)}}\sqrt{A_{ll}^{(m)}}} \le C\exp\left(-0.25\ b\,(U_\varphi + L_\varphi)^{-(\beta+1)}\ 2^{m\beta}\right). \tag{9.11}$$

In addition, if $b = 0$ and $m_1 \le m \le J-1$, then, as $n \to \infty$,

$$\|\mathbf{B}^{(m)}\| = O\left(2^{-m\alpha/2}\right), \qquad \mathbb{E}(\hat{c}_k^{(m)} - c_k^{(m)})^4 = O\left(n^{-2}2^{-2m\alpha}\right). \tag{9.12}$$

Proof of Lemma 8. First, note that, by (5.9), one has $\mathrm{Var}(\hat{c}_k^{(m)}) = n^{-1}\int \varphi_{mk}^2(x)(f^2(x) + \sigma^2)g(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)dx = O\left(n^{-1}\, A_{kk}^{(m)}\right)$. If $b = 0$, then

$$\begin{aligned} \mathbb{E}(\hat{c}_k^{(m)} - c_k^{(m)})^4 &= O\left(n^{-3}\int \varphi_{mk}^4(x)g(x)dx + n^{-2}\left[\int \varphi_{mk}^2(x)g(x)dx\right]^2\right) \\ &= O\left(n^{-3}\,2^m 2^{-m\alpha} + n^{-2}2^{-2m\alpha}\right) = O\left(n^{-2}2^{-2m\alpha}\right) \end{aligned}$$

since $n^{-1}2^{m(1+\alpha)} < 1$ for $m \le J-1$, which completes the proof of the second half of (9.12). Now, observe that, as $n \to \infty$,

$$\begin{aligned} A_{kl}^{(m)} &= \int \varphi(z + k_{0m} - k)\varphi(z + k_{0m} - l)g(x_0 + 2^{-m}z)\mathbb{I}(z + k_{0m} - l \in \Omega_\delta)dz \qquad (9.13) \\ &\sim C_g\ 2^{-m\alpha}\int \varphi(z + k_{0m} - k)\varphi(z + k_{0m} - l)|z|^\alpha\exp(-b2m\beta|z|^{-\beta})\,dz, \quad k, l \in K_{0m}^\varphi (9.14) \end{aligned}$$

and $B^{(m)}_{kl}$ has a similar expression, just with $k \in K^*_{0m}$ and $l \in K^{\varphi}_{0m}$, where K^*_{0m} is defined in (5.7). Recalling that $b = 0$ and the quantities $|k - k_{0m}|$ and $|l - k_{0m}|$ are uniformly bounded for $k \in K^{\varphi}_{0mc}$ and $l \in K^{\varphi}_{0m}$, obtain (for $b = 0$) that $|B^{(m)}_{kl}| = O(2^{-m\alpha})$, so that the first statement in (9.12) is true due to the fact that matrix $\mathbf{B}^{(m)}$ is finite dimensional.

Now, let $b > 0$ and let us prove (9.10). Observe that $L_\varphi \le z + k_{0m} - k \le U_\varphi$ and $k \in K^{\varphi}_{0m}$ imply $|z| \le M_\varphi$. Hence, the upper bound in (9.10) follows from (2.2) and (9.13). In order to prove the lower bound in (9.10), note that

$$A^{(m)}_{kk} \ge C_{g1} 2^{-m\alpha} \int_{\Omega^*_\delta} \varphi^2(z)|z - (k_{0m} - k)|^\alpha \exp(-b2m\beta|z - (k_{0m} - k)|^{-\beta})\, dz$$

where $\Omega^*_\delta = (L_\varphi + \delta_b, (L_\varphi + U_\varphi - 1)/2) \cup ((L_\varphi + U_\varphi + 1)/2, U_\varphi - \delta_b)$ and δ_b is defined in (5.6). Since $|z - (k_{0m} - k)| \ge 1/2$ for $z \in \Omega^*_\delta$, and by (4.2), $(L_\varphi + U_\varphi - 1)/2 - (L_\varphi + \delta_b) \ge 1$ and $(U_\varphi - \delta_b) - (L_\varphi + U_\varphi + 1)/2 \ge 1$, one has

$$A^{(m)}_{kk} \ge C_{g1} 2^{-\alpha(m+1)}\ \exp\left(-b2^{\beta(m+1)}\right)\ \min\left(\int_{L_\varphi+\delta_b}^{(L_\varphi+U_\varphi-1)/2} \varphi^2(z)dz,\ \ \int_{(L_\varphi+U_\varphi+1)/2}^{U_\varphi-\delta_b} \varphi^2(z)dz\right),$$

which completes the proof of (9.10).

Finally, let us prove (9.11). Note that asymptotic value of the integral in (9.13) is defined by the value at a point which maximizes the argument of the exponential function. Recall that (see, e.g., Dingle (1973)) if $F(\lambda) = \int_a^b h(x)\exp(\lambda S(x))dx$ where $\max S(x)$ is achieved at $x = a$ and $S(x)$ is a decreasing function of x, functions $f(x)$ and $S(x)$ are continuous on $[a, b]$ and infinitely differentiable in the neighborhood of $x = a$ with $S'(a) \ne 0$, then, as $\lambda \to \infty$, $F(\lambda)$ has the following asymptotic expression

$$F(\lambda) \sim \exp(\lambda S(a)) \sum_{k=o}^{\infty} c_k \lambda^{-(k+1)} \quad \text{with} \quad c_k = -D^k(h(x)/S'(x)) \tag{9.15}$$

where D is the differential operator of the form $D = -\frac{1}{S'(x)}\frac{d}{dx}$.

It is easy to calculate that $\exp(-b2m\beta|z|^{-\beta})$ takes its maximum value at $z = z^{(l,k,\delta)}_{\max} = \max(u_\delta, v_\delta$ where $u^{(l,k)}_\delta = \max(L_\varphi + k - k_{0m}, L_\varphi + \delta_b + l - k_{0m})$, and $v^{(l,k)}_\delta = \min(U_\varphi + k - k_{0m}, U_\varphi - \delta_b + l - k_{0m})$ and $L_\varphi \le k - k_{0m}, l - k_{0m} \le U_\varphi$. In what follows, we shall drop the superscripts whenever it does not cause confusion.

First, consider the case of $k = l$. Then, by examining the cases $k_{0m} - l \le (L_\varphi + U_\varphi)/2$ and $k_{0m} - l > (L_\varphi + U_\varphi)/2$ separately, one can easily conclude that

$$z^{(l,l,\delta)}_{\max} = \begin{cases} |L_\varphi + \delta_b + l - k_{0m}|, & \text{if} \quad k_{0m} - l > (L_\varphi + U_\varphi)/2, \\ |U_\varphi - \delta_b + l - k_{0m}|, & \text{if} \quad k_{0m} - l \le (L_\varphi + U_\varphi)/2, \end{cases} \tag{9.16}$$

where, by (4.2), $z^{(l,l,\delta)}_{\max} \ge (U_\varphi - L_\varphi - 2\delta_b)/2 \ge 2 - \delta_b > 1$ in both cases. Hence, since $\varphi(z^{(l,l,\delta)}_{\max}) \ne 0$ by definition of δ_b, formula (9.15) yields

$$\begin{aligned} A^{(m)}_{ll} &\sim C_g(b\beta)^{-1}\varphi^2(z^{(l,l,\delta)}_{\max} + k_{0m} - l)\ |z^{(l,l,\delta)}_{\max}|^\alpha\ 2^{-m(\alpha+\beta)}\ \exp(-b2^{m\beta}|z^{(l,l,\delta)}_{\max}|^{-\beta}) \\ &\ge K_1 2^{-m(\alpha+\beta)}\ \exp(-b2^{m\beta}|z^{(l,l,\delta)}_{\max}|^{-\beta}). \end{aligned} \tag{9.17}$$

If $k \ne l$, then $|k - l| \ge 1$ and one has four cases, depending on whether $k_{0m} - k$ and $k_{0m} - l$ are smaller or greater than $(L_\varphi + U_\varphi)/2$. We shall consider two of those since the other two cases are similar. In what follows, we denote by $z^{(k,k,0)}_{\max}$ the value of $z^{(k,k,\delta)}_{\max}$ obtained if $\delta = \delta_b = 0$.

If $k_{0m} - l \leq (L_\varphi + U_\varphi)/2$ and $k_{0m} - k \leq (L_\varphi + U_\varphi)/2$ then $|z_{\max}^{(l,l,\delta)}| = U_\varphi - \delta_b + l - k_{0m}$, $|z_{\max}^{(k,k,0)}| = U_\varphi + k - k_{0m}$ and, since $\delta_b < 1/2$,

$$|z_{\max}^{(l,k,\delta)}| = \begin{cases} U_\varphi - \delta_b + l - k_{0m}, & \text{if} \quad l > k \\ U_\varphi + k - k_{0m}, & \text{if} \quad l < k. \end{cases}$$

Therefore, taking into account that $|z_{\max}^{(k,k,\delta)}| = |z_{\max}^{(k,k,0)}| - \delta_b$, one derives that

$$\max\left(|z_{\max}^{(l,l,\delta)}|, |z_{\max}^{(k,k,\delta)}|\right) - |z_{\max}^{(l,k,\delta)}| \geq 1 - 2\delta_b. \tag{9.18}$$

Now, consider the case when $k_{0m} - l \leq (L_\varphi + U_\varphi)/2$ and $k_{0m} - k > (L_\varphi + U_\varphi)/2$. In this situation, $|z_{\max}^{(l,l,\delta)}| = U_\varphi - \delta_b + l - k_{0m}$, $|z_{\max}^{(k,k,0)}| = k_{0m} - k - L_\varphi$ and $|z_{\max}^{(l,k,\delta)}| = \max(|U_\varphi + k - k_{0m}|, |L_\varphi + \delta_b + l - k_{0m}|)$, so that relation (9.18) is again true. Cases when $k_{0m} - l > (L_\varphi + U_\varphi)/2$ can be examined in a similar manner and it can be shown that (9.18) is valid.

The asymptotic expression for $A_{kl}^{(m)}$ as $m \to \infty$ can be obtained using formula (9.15)

$$A_{kl}^{(m)} \sim C_g K(\varphi, b, \beta, z_{\max}) \; 2^{-m(\alpha+\beta)}\, 2^{-m\beta r^*} \exp(-b2m\beta|z_{\max}|^{-\beta}), \tag{9.19}$$

where $K(\varphi, b, \beta, z_{\max})$ depends on φ, b, β and $z_{\max}$ only and, hence, uniformly bounded, $r^* = 0$ if $z_{\max}$ does not coincide with L_φ or U_φ and $r^* = r_0 + 1$ if it does. Here, r_0 is the number of continuous derivatives of φ.

We are now ready to complete the proof of the lemma. Recall that

$$|z_{\max}^{(l,k,\delta)}| \leq \min(|z_{\max}^{(l,l,\delta)}|, |z_{\max}^{(k,k,0)}|) \leq \min(|z_{\max}^{(l,l,\delta)}|, |z_{\max}^{(k,k,\delta)}|) + \delta_b,$$

and, by (9.18), that

$$|z_{\max}^{(l,k,\delta)}| \leq \max(|z_{\max}^{(l,l,\delta)}|, |z_{\max}^{(k,k,\delta)}|) - (1 - 2\delta_b).$$

Since $|z_{\max}^{(l,l,\delta)}| > 2 - \delta_b$ and $|z_{\max}^{(k,k,\delta)}| > 2 - \delta_b$, an application of Lemma 7, with $a = |z_{\max}^{(l,l,\delta)}|$, $b = |z_{\max}^{(k,k,\delta)}|$, $c = |z_{\max}^{(l,k,\delta)}|$ and $M = (L_\varphi + U_\varphi)/2$, completes the proof of the lemma.

Lemma 9 *Let* $\mathbf{A}$ *be the matrix with the entries given by (5.9) and let* $\mathbf{D}$ *be the diagonal matrix* $\mathbf{D} = \sqrt{diag(\mathbf{A})}$. *Denote* $\mathbf{Q} = \mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}$. *Then, for any* $b \geq 0$ *one has* $\|\mathbf{Q}^{-1}\| = O(1)$ *as* $m \to \infty$. *Moreover, if* $b > 0$, *then* $\mathbf{Q}^{-1} = \mathbf{I} + \mathbf{H}$, *where*

$$\|\mathbf{H}\| = O\left(\exp(-0.125\ b\delta_0^2 2^{m\beta})\right), \quad m \to \infty,$$

and δ_0 *is defined in Lemma 8, i.e.,* $\mathbf{Q}^{-1} = \mathbf{I}\,(1 + o(1))$ *as* $m \to \infty$.

Proof of Lemma 9. Note that matrix $\mathbf{Q}$ is an $(U_\varphi - L_\varphi + 1)$-dimensional positive definite matrix with a unit main diagonal and smaller off-diagonal entries, so that, it has a non-asymptotic bounded inverse $\mathbf{Q}^{-1}$ with $\|\mathbf{Q}^{-1}\| = O(1)$.

If $b > 0$, then $Q_{kk} = 1$, so that $\mathbf{Q} = \mathbf{I} + \mathbf{H}$. Here, by Lemma 9, $\mathbf{H}$ is a finite dimensional matrix with elements $H_{lk} = O\left(\exp\left\{-0.25\, b\, (U_\varphi + L_\varphi)^{-(\beta+1)}\ 2^{m\beta}\right\}\right)$, as $m \to \infty$. Hence, $\|\mathbf{H}\| = O\left(\exp\left\{-0.25\, b\, (U_\varphi + L_\varphi)^{-(\beta+1)}\ 2^{m\beta}\right\}\right)$. To complete the proof of the lemma, it suffices to note that

$$\mathbf{Q}^{-1} = \mathbf{I} + \sum_{k=1}^{\infty} (-1)^k \mathbf{H}^k \quad \text{where} \quad \left\| \sum_{k=1}^{\infty} (-1)^k \mathbf{H}^k \right\| \leq \sum_{k=1}^{\infty} \|\mathbf{H}\|^k = O\left(\exp\left\{-\frac{kb\ 2^{m\beta}}{4\,(U_\varphi + L_\varphi)^{\beta+1}}\right\}\right).$$

9.4 Proofs of the large deviation results

Denote

$$\varrho_n = n^{-1/2}\sqrt{\ln n}. \tag{9.20}$$

In order to prove Lemma 4, we need the following three large deviation results.

Lemma 10 *Let $b = 0$. Let $w(x)$ be a bounded function with a bounded support (W_1, W_2) and a unit L^2-norm. Denote $w_{jk}(x) = 2^{j/2}w(2^j x - k)$ and set*

$$\beta_{jk} = \int w_{jk}(x)f(x)dx, \qquad \hat{\beta}_{jk} = n^{-1}\sum_{l=1}^{n}\frac{w_{jk}(x_i)y_i}{g(x_i)},$$

where f is the unknown response function in model (1.1). Let C_{g1} be defined in (2.2), and let

$$C_w = [2\max(|W_1|,|W_2|)]^{\alpha}, \quad C_\tau = 8C_w\,(C_{g1})^{-1}\ \max\left(2,\ 2\|f\|_\infty^2, \|f\|_\infty\|w\|_\infty/3, \|w\|_\infty\right). \tag{9.21}$$

Then, for $m_1 \le j \le J-1$, $k \in K_{0jc}^w$ and $\tau \ge 1$, as $n \to \infty$, one has

$$\mathbb{P}\left(|\hat{\beta}_{jk} - \beta_{jk}| > \tau\,\varrho_n\,2^{j\alpha/2}|k - k_{0j}|^{-\alpha/2}\right) = O\left(n^{-\frac{\tau}{C_\tau}}\right). \tag{9.22}$$

Here, m_1 and J are defined in (5.3), ϱ_n is defined (9.20), and

$$K_{0jc}^w = \left\{k:\ 0 \le k \le 2^j - 1,\ x_0 \notin supp\ w_{jk}\right\}.$$

Proof of Lemma 10. First, note that a slightly unusual formulation of this lemma is due to the fact that we are planning to use it both with $w = \varphi$ and $w = \psi$. The proof of the lemma is based on ideas presented in Chesneau (2007a). Observe that

$$\mathbb{P}\left(|\hat{\beta}_{jk} - \beta_{jk}| > \tau\,\varrho_n\,2^{j\alpha/2}|k - k_{0j}|^{-\alpha/2}\right) \le P_1 + P_2,$$

where

$$\begin{aligned}
P_1 &= \mathbb{P}\left(\left|n^{-1}\ \sum_{i=1}^{n}[g(x_i)]^{-1}\,w_{jk}(x_i)f(x_i) - \beta_{jk}\right| > 0.5\,\tau\,\varrho_n\,2^{j\alpha/2}|k - k_{0j}|^{-\alpha/2}\right),\\
P_2 &= \mathbb{P}\left(\left|n^{-1}\ \sum_{i=1}^{n}[g(x_i)]^{-1}\,w_{jk}(x_i)\xi_i\right| > 0.5\,\tau\,\varrho_n\,2^{j\alpha/2}|k - k_{0j}|^{-\alpha/2}\right).
\end{aligned}$$

The proof of the statement is based on Bernstein inequality

$$\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^{n}\eta_i\right| > z\right) \le 2\ \exp\left(-\frac{nz^2}{2(\sigma^2 + \|\eta\|_\infty\,z/3)}\right), \tag{9.23}$$

where $\eta_i,\ i = 1, 2, \ldots, n$, are i.i.d. with $\mathbf{E}\eta_i = 0$, $\mathbf{E}\eta_i = \sigma^2$ and $\|\eta_i\| \le \|\eta\|_\infty < \infty$.

First, let us construct an upper bound for P_1. Note that for $k \in K_{0jc}^w$ one has

$$g(x_i)\mathbb{I}(x_i \in \operatorname{supp} w_{jk}) \ge C_w^{-1}\,C_{g1}2^{-j\alpha}|k - k_{0j}|^{\alpha}. \tag{9.24}$$

Let $\eta_i = [g(x_i)]^{-1}\, w_{jk}(x_i) f(x_i) - \beta_{jk}$. Then, $\mathbb{E}\eta_i = 0$ and, by (9.24), we derive $\|\eta\|_\infty \leq C_{g1}^{-1} C_w\, |k - k_{0j}|^{-\alpha}\, 2^{j(\alpha+1/2)}\, \|w\|_\infty \|f\|_\infty$, so that

$$\mathrm{Var}\eta_i = \int \frac{w_{jk}^2(x) f^2(x)}{g(x)}\, dx \leq \frac{\|f\|_\infty^2}{C_{g1}} \int_{W_1}^{W_2} \frac{w^2(t) 2^{j\alpha}}{|t + k - k_{0j}|^\alpha} dt \leq \frac{\|f\|_\infty^2\, 2^{j\alpha} C_w}{C_{g1}\, |k - k_{0j}|^\alpha}.$$

Now, applying Bernstein's inequality and recalling that $j \leq J$, $2^{J(\alpha+1)} = n/\ln n$ and $|k - k_{0j}| \geq 1$, we obtain

$$P_1 \quad \leq \quad 2 \exp\left(- \frac{C_{g1}\, \tau^2 \ln n}{8 C_w\, \|f\|_\infty (\|f\|_\infty + \|w\|_\infty\, \tau/6)} \right).$$

Using the inequality $a/(b+c) \geq \min\left(a/(2b), a/(2c)\right)$, where $a, b, c > 0$, and taking into account that $\tau^2 \geq \tau$ for $\tau \geq 1$, we obtain

$$P_1 \leq 2 \exp(-\tau \ln n / D_1) \quad \text{with} \quad D_1 = 8\, C_{g1}^{-1} C_w\, \max(2\|f\|_\infty^2, \|f\|_\infty \|w\|_\infty / 3). \tag{9.25}$$

In order to construct an upper bound for P_2, note that, conditionally on $(x_1, x_2, \ldots, x_n)$, one has

$$n^{-1} \sum_{i=1}^n (g(x_i))^{-1}\, w_{jk}(x_i) \xi_i \sim \mathcal{N}(0, s_{jk}^2),$$

where, by (9.24) and $\sigma = 1$,

$$s_{jk}^2 = \frac{1}{n^2} \sum_{i=1}^n \frac{w_{jk}^2(x_i)}{g^2(x_i)} \leq \frac{C_w 2^{j\alpha}}{C_{g1} |k - k_{0j}|^\alpha\, n^2} \sum_{i=1}^n \frac{w_{jk}^2(x_i)}{g(x_i)}.$$

Hence, conditionally on $(x_1, x_2, \ldots, x_n)$,

$$\mathbb{P}\left(\left| n^{-1} \sum_{i=1}^n [g(x_i)]^{-1}\, w_{jk}(x_i)\, \xi_i \right| > \frac{\tau\, \varrho_n\, 2^{j\alpha/2}}{2\, |k - k_{0j}|^{\alpha/2}} \middle| x_1, x_2, \ldots, x_n \right) \leq \exp\left(- \frac{\tau^2 2^{j\alpha} \ln n}{8 n |k - k_{0j}|^\alpha s_{jk}^2} \right).$$

Now, consider the following two sets:

$$\Omega_\upsilon(x_1, x_2, \ldots, x_n) = \left\{ (x_1, x_2, \ldots, x_n) : \left| \frac{1}{n} \sum_{i=1}^n \frac{w_{jk}^2(x_i)}{g(x_i)} - 1 \right| \geq \upsilon \right\},$$

and its complementary, $\Omega_\upsilon^c(x_1, x_2, \ldots, x_n)$. Then $P_2 \leq P_{21} + P_{22}$ where

$$\begin{aligned} P_{21} &= \mathbb{E}\left[\mathbb{P}\left(\left| n^{-1} \sum_{i=1}^n (g(x_i))^{-1}\, w_{jk}(x_i) \xi_i \right| > \frac{\tau\, \varrho_n\, 2^{j\alpha/2}}{2\, |k - k_{0j}|^{\alpha/2}} \middle| x_1, x_2, \ldots, x_n \right) \mathbb{I}(\Omega_\upsilon^c(x_1, x_2, \ldots, x_n)) \right], \\ P_{22} &= \mathbb{E}\left[\mathbb{I}(\Omega_\upsilon(x_1, x_2, \ldots, x_n)) \right] \end{aligned}$$

and $\mathbb{I}(\Omega)$ is the indicator of the set Ω. Since for $\Omega_\upsilon^c(x_1, x_2, \ldots, x_n)$, we have

$$n^{-1} \sum_{i=1}^n [g(x_i)]^{-1}\, w_{jk}^2(x_i) \leq \left(1 + n^{-1} \left| \sum_{i=1}^n [g(x_i)]^{-1}\, w_{jk}^2(x_i) - 1 \right| \right) \leq \upsilon + 1,$$

and it is easy to check that

$$P_{21} \leq \exp(-\tau^2 \ln n / D_2) \quad \text{with} \quad D_2 = 8\, C_{g1}^{-1} C_w\, (\upsilon + 1). \tag{9.26}$$

In order to find an upper bound for P_{22}, we apply Bernstein's inequality with $Z_i = [g(x_i)]^{-1} w_{jk}^2(x_i)$. Note that $\mathrm{Var} Z_i \leq C_{g1}^{-1} C_w \|w\|_\infty 2^{j(\alpha+1)} |k - k_{0j}|^{-\alpha}$, $\|Z\|_\infty \leq 2C_{g1}^{-1} C_w \|w\|_\infty 2^{j(\alpha+1)} |k - k_{0j}|^{-\alpha}$ and $\mathbb{E} Z_i = 0$. Application of (9.23) with $z = \upsilon$, yields

$$P_{22} \leq 2 \exp(-\upsilon^2 \ln n / D_3) \quad \text{with} \quad D_3 = 2\,\|w\|_\infty^2\, C_{g1}^{-1} C_w \,(1 + 2\upsilon/3). \tag{9.27}$$

Now, set $\upsilon = 0.5\,\tau \|w\|_\infty$ and observe that for this value of υ and $\tau \geq 1$, one has

$$4\|w\|_\infty^{-2}\,(1 + 2\upsilon/3)^{-1}\,\upsilon^2 \geq \tau^2/(\upsilon + 1) \geq \tau \cdot \min(1/2, \|w\|_\infty^{-1}).$$

To complete the proof, we only need to combine (9.25), (9.26) and (9.27).

Lemma 11 *Let $b = 0$, $C_\varphi = [2 \max(|L_\varphi|, |U_\varphi|)]^\alpha$ and C_{g2} be defined in (2.2). Let m be a non-random integer, $m_1 \leq m \leq J - 1$, and $k \in K_{0m}^{\varphi}$. Then, for $c_l^{(m)}$ and $\hat{c}_l^{(m)}$ given by (5.11) and (5.14), respectively, and an arbitrary constant $\kappa \geq 1$, one has*

$$\mathbb{P}\left(|\hat{c}_l^{(m)} - c_l^{(m)}| > \kappa\, \varrho_n\, 2^{-\frac{m\alpha}{2}} \right) = O\left(n^{-\frac{\kappa}{C_\kappa}} \right), \quad n \to \infty. \tag{9.28}$$

Here, ϱ_n is defined in (9.20) and C_κ is given by formula (6.9).

Proof of Lemma 11. The proof is very similar to the proof of Lemma 10, therefore, we shall just provide its outline. Partition the probability in (9.28) into P_1 and P_2 with

$$\begin{aligned} P_1 &= \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^n \varphi_{mk}(x_i) f(x_i) - c_k^{(m)} \right| > 0.5\,\kappa\, \varrho_n\, 2^{-\frac{m\alpha}{2}} \right), \\ P_2 &= \mathbb{P}\left(\left| n^{-1} \sum_{i=1}^n \varphi_{mk}(x_i) \xi_i \right| > 0.5\,\kappa\, \varrho_n\, 2^{-\frac{m\alpha}{2}} \right). \end{aligned}$$

An upper bound for P_1, obtained by applying Bernstein's inequality, is of the form

$$P_1 \leq 2 \exp(-\kappa \ln n / D_4) \quad \text{with} \quad D_4 = 8\,\|f\|_\infty\, \max(2C_{g2} C_\varphi, \|\varphi\|_\infty/3). \tag{9.29}$$

In order to derive an upper bound for P_2, introduce a set

$$\Theta_v(x_1, x_2, \ldots, x_n) = \left\{ (x_1, x_2, \ldots, x_n) : \left| \frac{1}{n} \sum_{i=1}^n \varphi_{mk}^2(x_i) - \int \varphi_{mk}^2(x) g(x) dx \right| \geq v 2^{-m\alpha} \right\},$$

and its complementary, $\Theta_v^c(x_1, x_2, \ldots, x_n)$. Then, similarly to the proof of Lemma 10, obtain $P_2 \leq P_{21} + P_{22}$, where

$$\begin{aligned} P_{21} &= \mathbb{E}\left[\mathbb{P}\left(\left| n^{-1} \sum_{i=1}^n \varphi_{mk}(x_i) \xi_i \right| > 0.5\,\kappa\, \varrho_n\, 2^{-m\alpha/2} \middle| x_1, x_2, \ldots, x_n \right) \mathbb{I}(\Theta_v^c(x_1, x_2, \ldots, x_n)) \right] \\ &\leq \exp\left(-\frac{\kappa^2 \ln n}{8(v + C_\varphi C_{g2})} \right). \end{aligned}$$

Also, application of (9.23) with $\eta_i = \left(\varphi_{mk}^2(x_i) - \int \varphi_{mk}^2(x) g(x) dx \right)$, yields

$$P_{22} = \mathbb{E}\left[\mathbb{I}(\Theta_v^c(x_1, x_2, \ldots, x_n)) \right] \leq 2 \exp\left(-\frac{n v^2\, 2^{-m(1+\alpha)}}{2\|\varphi\|_\infty^2 (C_\varphi C_{g2} + v/3)} \right).$$

Setting $v = a\kappa$, noting that for any $A, B, C > 0$ one has $A/(b+c) \ge \min(A/(2B), A/(2C))$, and recalling that $\kappa \ge 1$ and $n2^{-m(1+\alpha)} \ge \ln n$ by (5.3), we derive

$$P_2 \le 2\exp(-\kappa \ln n/D_5) \quad \text{with} \quad D_5 = \max\left(16a, 16C_\varphi C_{g2}, \frac{4C_\varphi C_{g2}\|\varphi\|_\infty}{a}, \frac{4\|\varphi\|_\infty^2}{3a}\right). \tag{9.30}$$

To complete the proof, it suffices to note that $a > 0$ is arbitrary.

Lemma 12 *Let $b = 0$, let m_0 and $\hat{m}$ be given by (5.18) and (6.2), respectively. Consider the non-asymptotic finite dimension matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ with elements*

$$\begin{aligned} A^*_{kl} &= \int \varphi(z + k_{0m} - k)\varphi(z + k_{0m} - l)|z|^\alpha dz, \quad k, l \in K^\varphi_{0m}, & (9.31)\\ B^*_{lk} &= \int \varphi(z + k_{0m} - k)\varphi(z + k_{0m} - l)|z|^\alpha dz, \quad l \in K^\varphi_{0m},\ k \in K^*_{0m}. & (9.32)\end{aligned}$$

Let $\hat{\mathbf{u}}^{(m)}$ be the solution of the system of equations (5.15).

If $\lambda \ge \max(C_{\lambda 1}, C_{\lambda 2})$, then, as $n \to \infty$,

$$\mathbb{P}\left(\|\hat{\mathbf{u}}^{(m)} - \mathbb{E}\hat{\mathbf{u}}^{(m)}\| > \lambda\, \varrho_n\, 2^{m\alpha/2}\right) = O\left(n^{-\frac{2\lambda}{C_u}}\right), \tag{9.33}$$

where $C_{\lambda 1}$ and $C_{\lambda 2}$ are defined in (6.5) and C_u is defined in (6.7).

Proof of Lemma 12. Observe that for any m, by (5.15), one has

$$\|\hat{\mathbf{u}}^{(m)} - \mathbb{E}\hat{\mathbf{u}}^{(m)}\| \le \|(\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)})\| + \|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}(\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)})\|,$$

so that

$$\begin{aligned} &\mathbb{P}\left(\|\hat{\mathbf{u}}^{(m)} - \mathbb{E}\hat{\mathbf{u}}^{(m)}\| > \lambda\, \varrho_n\, 2^{m\alpha/2}\right) \le \mathbb{P}\left(\|(\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)})\| > 0.5\,\lambda\, \varrho_n\, 2^{m\alpha/2}\right)\\ &+ \mathbb{P}\left(\|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}(\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)})\| > 0.5\,\lambda\, \varrho_n\, 2^{m\alpha/2}\right) \equiv P_1 + P_2.\end{aligned}$$

Now note that, by assumption (2.1) and the dominated convergence theorem, as $n \to \infty$, one has $\mathbf{A}^{(m)} = C_g 2^{-m\alpha}\mathbf{A}^*(1 + o(1))$ and $\mathbf{B}^{(m)} = C_g 2^{-m\alpha}\mathbf{B}^*(1 + o(1))$, where the matrices $\mathbf{A}^*$ and $\mathbf{B}^*$, defined in (9.31) and (9.32), are independent of m, since the sets K^φ_{0m} and K^*_{0m} are defined in terms of $k - k_{0m}$ and $l - k_{0m}$. Therefore, $\|(\mathbf{A}^{(m)})^{-1}\| = C_g^{-1}\, 2^{m\alpha}\, \|(\mathbf{A}^*)^{-1}\|(1 + o(1))$ and $\|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}\| = \|(\mathbf{A}^*)^{-1}\, \mathbf{B}^*\|(1 + o(1))$. Hence, setting $\kappa = C_{\lambda 1}^{-1}\, \lambda$ in Lemma 11, where $C_{\lambda 1}$ is defined in (6.5), and taking into account that the set K^φ_{0m} contains no more than $U_\varphi - L_\varphi + 1$ indices, we obtain

$$\begin{aligned} P_1 &\le \mathbb{P}\left(\|\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)}\| > \frac{C_{g2}\,\lambda\,\varrho_n\, 2^{-m\alpha/2}}{2\,\|(\mathbf{A}^*)^{-1}\|}\right) \le \sum_{k \in K^\varphi_{0m}} \mathbb{P}\left(|\hat{c}^{(m)}_k - c^{(m)}_k| > \frac{C_{g2}\,\lambda\,\varrho_n\, 2^{-m\alpha/2}}{2\sqrt{U_\varphi - L_\varphi + 1}\;\|(\mathbf{A}^*)^{-1}\|}\right)\\ &= \sum_{k \in K^\varphi_{0m}} \mathbb{P}\left(|\hat{c}^{(m)}_k - c^{(m)}_k| > 2C_{\lambda 1}^{-1}\,\lambda\,\varrho_n\, 2^{-m\alpha/2}\right) = O\left(n^{-2(C_\kappa\, C_{\lambda 1})^{-1}\,\lambda}\right).\end{aligned}$$

Similarly, using Lemma 10 with $w = \varphi$ and C_τ given by (6.8), and recalling the definitions of $\hat{\mathbf{v}}^{(m)}$ and $\mathbf{v}^{(m)}$, one can derive an upper bound for P_2 as

$$\begin{aligned} P_2 &\le \mathbb{P}\left(\|\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)}\| > \frac{\lambda\,\varrho_n\, 2^{m\alpha/2}}{2\,\|(\mathbf{A}^*)^{-1}\,\mathbf{B}^*\|}\right) \le \sum_{k \in K^*_{0m}} \mathbb{P}\left(|\hat{a}_{mk} - a_{mk}| > \frac{\lambda\,\varrho_n\, 2^{m\alpha/2}}{2\sqrt{U_\varphi - L_\varphi + 1}\;\|(\mathbf{A}^*)^{-1}\,\mathbf{B}^*\|}\right)\\ &= O\left(n^{-2(C_\tau\, C_{\lambda 2})^{-1}\,\lambda}\right),\end{aligned}$$

which completes the proof of the lemma.

Proof of Lemma 4. Note that by definition of $\hat{m}$, whenever $\hat{m} > m_0$, there exists $j > m_0$ such that $\|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Xi_{m_0})\|^2 > \lambda^2\, 2^{j\alpha}\, \rho_n^2$, where ϱ_n is defined in (9.20). Therefore,

$$\mathbb{P}(\hat{m} > m_0) \quad \le \quad \sum_{j=m_0}^{J-1} \mathcal{P}_j \ \text{ with } \ \mathcal{P}_j = \mathbb{P}\left(\|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Xi_{m_0})\|^2 > \lambda^2\, 2^{j\alpha}\, \rho_n^2\right). \tag{9.34}$$

Observe that since

$$\begin{aligned} \|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Xi_{m_0})\| \quad &\le \quad \|(\hat{f}_{0,j} - f_{0,j})\mathbb{I}(\Xi_{m_0})\| + \|(\hat{f}_{c,j} - f_{c,j})\mathbb{I}(\Xi_{m_0})\| \\ &+ \quad \|(\hat{f}_{0,m_0} - f_{0,m_0})\mathbb{I}(\Xi_{m_0})\| + \|(\hat{f}_{c,m_0} - f_{c,m_0})\mathbb{I}(\Xi_{m_0})\|, \end{aligned}$$

one has the following upper bound for $\mathcal{P}_j$ defined in (9.34):

$$\mathcal{P}_j \quad \le \quad \mathcal{P}_{0,j,m_0} + \mathcal{P}_{0,j,j} + \mathcal{P}_{c,j,m_0} + \mathcal{P}_{c,j,j}$$

where, for any $m_0 \le m \le j$,

$$\begin{aligned} \mathcal{P}_{0,j,m} \quad &= \quad \mathbb{P}\left(\|(\hat{f}_{0,m} - f_{0,m})\mathbb{I}(\Xi_{m_0})\| > 0.25\,\lambda\, 2^{j\alpha/2}\, \rho_n\right), \\ \mathcal{P}_{c,j,m} \quad &= \quad \mathbb{P}\left(\|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(\Xi_{m_0})\| > 0.25\,\lambda\, 2^{j\alpha/2}\, \rho_n\right). \end{aligned}$$

Since $\text{supp}(f_{0,m}) \subseteq \Xi_m \in \Xi_{m_0}$ for $m \ge m_0$, one has

$$\begin{aligned} \|(\hat{f}_{0,m} - f_{0,m})\mathbb{I}(\Xi_{m_0})\|^2 \quad &= \quad \|(\hat{f}_{0,m} - f_{0,m})\mathbb{I}(\Xi_m)\|^2 = \|\hat{f}_{0,m} - f_{0,m}\|^2 \\ &\le \quad \|\hat{\mathbf{u}}^{(m)} - \mathbf{u}^{(m)}\|^2 + 2(U_\varphi - L_\varphi + 1)A^2 2^{-2ms'}. \end{aligned} \tag{9.35}$$

Hence, by (9.35) and Lemma 12, since $m_0 \le m \le j$, one derives

$$\mathcal{P}_{0,j,m} \quad \le \quad \mathbb{P}\left(\|\hat{\mathbf{u}}^{(m)} - \mathbf{u}^{(m)}\| > 0.25\,\lambda\, 2^{j\alpha/2}\, \rho_n - A\sqrt{2(U_\varphi - L_\varphi + 1)}2^{-js'}\right) = O\left(n^{-\frac{\lambda}{2C_u}}\right).$$

Now, let us consider the second term, $\mathcal{P}_{c,j,m}$. Note that $\text{supp}(\varphi_{mk})$ and Ξ_{m_0} have non-empty intersection if and only if $k \in \tilde{K}_{m,m_0}$, where

$$\tilde{K}_{m,m_0} = \left\{k :\ 2^{m-m_0}[\min(L_\varphi, L_\psi) - U_\varphi] - U_\varphi < k - k_{0m} < 2^{m-m_0}[\max(U_\varphi, U_\psi) - L_\varphi] - L_\varphi\right\}.$$

Hence, for $m \ge m_0$,

$$\|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(\Xi_{m_0})\|^2 \quad \le \quad \|\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)}\|^2 + \sum_{j'=m}^{J-1} \sum_{k \in \tilde{K}_{j',m_0}} (\hat{b}_{j'k} - b_{j'k})^2 + \sum_{j=J}^{\infty} \sum_{k \in \tilde{K}_{j,m_0}} b_{jk}^2.$$

Here, by (9.57), we have

$$\sum_{j=J}^{\infty} \sum_{k \in \tilde{K}_{j,m_0}} b_{jk}^2 \le A^2 2^{-2Js^*} = O\left(n^{-\frac{2s'}{2s'+\alpha}}\ (\ln n)^{\frac{2s'}{2s'+\alpha}}\right),$$

where s^* is defined in (9.56). Also,

$$(\hat{b}_{j'k} - b_{j'k})^2 \quad \le \quad (\hat{b}_{j'k} - b_{j'k})^2 \mathbb{I}(|\hat{b}_{j'k} - b_{j'k}| > 0.5d2^{j\alpha/2}|k - k_{0j'}|^{-\alpha/2}) + b_{j'k}^2$$

since $\mathbb{I}(|\hat{b}_{j'k}| > d2^{j\alpha/2}|k-k_{0j'}|^{-\alpha/2}) \leq \mathbb{I}(|b_{j'k}| > 0.5d2^{j\alpha/2}|k-k_{0j'}|^{-\alpha/2}) + \mathbb{I}(|\hat{b}_{j'k}-b_{j'k}| > 0.5d2^{j\alpha/2}|k-k_{0j'}|^{-\alpha/2})$ and, for $j \geq m_0$ and n large enough, $\mathbb{I}(|b_{j'k}| > 0.5d2^{j\alpha/2}|k-k_{0j'}|^{-\alpha/2}) = 0$. Denote $C_{LU} = \max(|\min(L_\varphi, L_\psi) - 2U_\varphi|, |\max(U_\varphi, U_\psi) - L_\varphi|)$ and observe that $\tilde{K}_{j',m_0} \subset \left\{k: \ |k-k_{0j'}| < 2^{j'-m_0} C_{LU}\right\}$. Hence, using Cauchy inequality and (9.56), one obtains

$$\sum_{j'=m}^{J-1} \sum_{k \in \tilde{K}_{j',m_0}} b_{j'k}^2 \leq A^2 (2C_{LU})^{(1-2/p)_+} \ 2^{-2m_0 s'} 2^{-2s^*(m-m_0)}.$$

Combining all inequalities above, we derive that for any $m \geq m_0$,

$$\begin{aligned} \|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(\Xi_{m_0})\|^2 \ &\leq \ \|\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)}\|^2 + A^2 2^{-2Js^*} + A^2(2C_{LU})^{(1-2/p)_+} \ 2^{-2m_0 s'} 2^{-2s^*(m-m_0)} \\ &+ \ \sum_{j'=m}^{J-1} \sum_{k \in \tilde{K}_{j',m_0}} (\hat{b}_{j'k} - b_{j'k})^2 \mathbb{I}(|\hat{b}_{j'k} - b_{j'k}| > 0.5d2^{j\alpha/2}|k-k_{0j'}|^{-\alpha/2}). \end{aligned}$$

Now, by Lemma 10 with $w = \varphi$ and $w = \psi$, obtain

$$\begin{aligned} \mathcal{P}_{c,j,m} \ &\leq \ \mathbb{P}\left(\|\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)}\| > 0.25\,\lambda\, 2^{j\alpha/2}\,\rho_n - A^2 2^{-2Js^*} + A^2(2C_{LU})^{(1-2/p)_+} \ 2^{-2m_0 s'} 2^{-2s^*(m-m_0)}\right) \\ &+ \ \sum_{j'=m}^{J-1} \sum_{k \in \tilde{K}_{j',m_0}} \mathbb{P}(|\hat{b}_{j'k} - b_{j'k}| > 0.5d2^{j\alpha/2}|k-k_{0j'}|^{-\alpha/2}) \\ &= \ O\left(n^{-(C_\tau C_{\lambda 0})^{-1}\lambda}\right) + O\left(n^{\frac{1}{\alpha+1} - \frac{d}{2C_d}}\right), \end{aligned}$$

which completes the proof.

9.5 Proofs of the statements in Section 5: risk of the zero-affected part of the wavelet estimator

Proof of Lemma 1. Note that

$$\Delta_1 = \|\mathbb{E}\hat{f}_0^{(m)} - f_0^{(m)}\|^2 = \sum_{j=m}^{\infty} \sum_{k \in K_{0m}^\varphi} b_{jk}^2, \quad \Delta_2 = \mathbb{E}\|\hat{f}_0^{(m)} - \mathbb{E}\hat{f}_0^{(m)}\|^2 = \sum_{k \in K_{0m}^\varphi} \mathbb{E}(\hat{a}_{mk} - a_{mk})^2,$$

where $\hat{a}_{mk} = \hat{u}_k^{(m)}$ for $k \in K_{0m}^\varphi$. From the characterization (3.1) of Besov spaces, it follows that, for any k, one has $b_{jk}^2 \leq A 2^{-2js'}$, and, therefore, since the number of indices in the set K_{0m}^φ is finite,

$$\Delta_1 \ = \ O\left(\sum_{j=m}^{\infty} 2^{-2js'}\right) = O\left(2^{-2ms'}\right). \tag{9.36}$$

Now, consider Δ_2. Let, as in Lemma 9, $\mathbf{A}^{(m)}$ be the matrix with the entries given by (5.9), $\mathbf{D}^{(m)} = \sqrt{\mathrm{diag}(\mathbf{A}^{(m)})}$ and $\mathbf{Q}^{(m)} = (\mathbf{D}^{(m)})^{-1}\mathbf{A}^{(m)}(\mathbf{D}^{(m)})^{-1}$. In the following proof, for the sake of clarity, we shall suppress the index m. Rewrite the systems of equations (5.8) and (5.15), respectively, as

$$\mathbf{Q}\,\mathbf{D}\,\mathbf{u} = \mathbf{D}^{-1}\,\mathbf{c} + \mathbf{D}^{-1}\,\boldsymbol{\varepsilon} - \mathbf{D}^{-1}\,\mathbf{B}\mathbf{v}, \quad \mathbf{Q}\,\mathbf{D}\,\hat{\mathbf{u}} = \mathbf{D}^{-1}\,\hat{\mathbf{c}} - \mathbf{D}^{-1}\,\mathbf{B}\hat{\mathbf{v}}, \tag{9.37}$$

so that

$$\hat{\mathbf{u}} - \mathbf{u} = \mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}(\hat{\mathbf{c}} - \mathbf{c}) - \mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v}) + \mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}\boldsymbol{\varepsilon}. \tag{9.38}$$

Therefore,

$$\Delta_2 \quad = \quad \mathbb{E}\|\hat{\mathbf{u}} - \mathbf{u}\|^2 = O\left(\Delta_{21} + \Delta_{22} + \Delta_{23}\right), \tag{9.39}$$

with

$$\Delta_{21} = \mathbb{E}\|\mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}(\hat{\mathbf{c}} - \mathbf{c})\|^2, \quad \Delta_{22} = \mathbb{E}\|\mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v})\|^2, \quad \Delta_{23} = \|\mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^2. \tag{9.40}$$

By Lemma 8, one has

$$\mathbf{D}_{ii} \geq C 2^{-m\alpha/2} \exp(-0.5\, b 2^{\beta(m+1)}),$$

and since $\mathbf{D}$ is the finite-dimensional diagonal matrix, the latter implies

$$\|\mathbf{D}^{-1}\| = O\left(2^{m\alpha/2} \exp(0.5\, b 2^{\beta(m+1)})\right). \tag{9.41}$$

Therefore, since the set K_{0m}^{φ} is finite, by Lemma 8, one has

$$\mathbb{E}\|\mathbf{D}^{-1}(\hat{\mathbf{c}} - \mathbf{c})\|^2 = \sum_{k \in K_{0m}^{\varphi}} \mathrm{Var}(\hat{c}_k^{(m)})/A_{kk}^{(m)} = O\left(n^{-1}\right),$$

so that we derive

$$\Delta_{21} = O\left(\|\mathbf{D}^{-1}\|^2 \|\mathbf{Q}^{-1}\|^2 \ \mathbb{E}\|\mathbf{D}^{-1}(\hat{\mathbf{c}} - \mathbf{c})\|^2\right) = O\left(n^{-1}\ 2^{m\alpha}\ \exp(b 2^{\beta(m+1)})\right). \tag{9.42}$$

In order to derive an upper bound for Δ_{22}, note that from (9.10), (9.40) and considerations above, it follows that

$$\Delta_{22} = O\left(\|\mathbf{D}^{-1}\|^4 \|\mathbf{Q}^{-1}\|^2 \ \mathbb{E}\|\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v})\|^2\right) = O\left(2^{2m\alpha} \exp(2b\, 2^{\beta(m+1)})\ \mathbb{E}\|\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v})\|^2\right).$$

Since $\exp(-b 2^{m\beta}|z|^{-\beta})$ is an increasing function of $|z|$ and, for $L_\varphi \leq z + k_{0m} - k \leq U_\varphi$ and $k \in K_{0m}^*$, one has $|z| \leq 2(U_\varphi - L_\varphi)$, for $k \in K_{0m}^*$, we derive

$$\begin{aligned} C_{kk} = \int \varphi_{mk}^2(x) g(x) \mathbb{I}(2^m x - l \in \Omega_\delta) dx \quad &= \quad O\left(2^{-m\alpha} \int \varphi^2(z + k_{0m} - k)\, |z|^\alpha \ \exp\left\{-b 2^{m\beta} |z|^{-\beta}\right\} dz\right) \\ &= \quad O\left(2^{-m\alpha}\ \exp\left\{-b 2^{m(\beta-1)} (U_\varphi - L_\varphi)^{-\beta}\right\}\right). \end{aligned}$$

Hence, since the sets K_{0m}^{φ} and K_{0m}^* are finite, by definition of vector $\hat{\mathbf{v}}$, Lemmas 6 and 8 and Cauchy inequality, we obtain

$$\begin{aligned} \mathbb{E}\|\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v})\|^2 \quad &= \quad \sum_{h \in K_{0m}^{\varphi}} \ \sum_{k,l \in K_{0m}^*} B_{hk} B_{hl} \mathrm{Cov}(\hat{a}_k, \hat{a}_l) \leq n^{-1} \sum_{h \in K_{0m}^{\varphi}} \ \sum_{k,l \in K_{0m}^*} J_{mkl} A_{hh} \sqrt{C_{kk} C_{ll}} \\ &= \quad O\left(n^{-1} 2^{-m\alpha} \exp\left\{b 2^{m\beta} - b 2^{m(\beta-1)} (U_\varphi - L_\varphi)^{-\beta} - b 2^{m\beta}\, M_\varphi^{-\beta}\right\}\right), \end{aligned}$$

where M_φ is defined in Lemma 8. Since $U_\varphi - L_\varphi \geq 4$, we finally obtain

$$\Delta_{22} = O\left(n^{-1}\ 2^{m\alpha}\ \exp(b\, 2^{m\beta} [2^{\beta+1} + 1])\right). \tag{9.43}$$

Now, for the function $\varepsilon_m(x)$ defined in (5.4), one has

$$\Delta_{23} = O\left(\|\mathbf{D}^{-1}\|^2\, \|\mathbf{Q}^{-1}\|^2\, \|\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^2\right) \tag{9.44}$$

where

$$\|\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^2 = \sum_{k \in K_{0m}^{\varphi}} A_{kk}^{-2} \left[\int \varepsilon_m(x)\, \varphi_{mk}^*(x)\, g(x) dx \right]^2. \tag{9.45}$$

If $b = 0$, by Cauchy-Schwarz inequality, one obtains $\|\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^2 \le \sum_{k \in K_{0m}^{\varphi}} A_{kk}^{-2} \|\varepsilon_{m0}\|^2 \;\|\varphi_{mk}\, g\|^2$, where $\varepsilon_{m0}(x) = \varepsilon_m(x)\mathbb{I}(|x - x_0| \le C2^{-m})$. By calculations similar to proof of Lemma 8 in the case of $b = 0$, one can show that $\|\varphi_{mk}\, g\|^2 = \int \varphi_{mk}^2(x) g(x) dx = O\left(2^{-2m\alpha}\right)$. Also, since $b_{jk}^2 \le A 2^{-2js'}$, one has

$$\begin{aligned} \|\varepsilon_{m0}\|^2 &= O\left(\sum_{j=m}^{\infty} \sum_{|k - k_{0j}| \le C2^{j-m}} \right) \\ &= O\left(\sum_{j=m}^{\infty} 2^{-2js'} 2^{(j-m)(1-2/p)} \right) = O\left(2^{-2ms'}\right). \end{aligned}$$

Recalling (9.10), we obtain in the case of $b = 0$,

$$\Delta_{23} = O\left(2^{-2ms'}\right). \tag{9.46}$$

Now, let us consider the case of $b > 0$. Denote $\varphi_{mk}^*(x) = \varphi_{mk}(x)\mathbb{I}(2^m x - k \in \Omega_\delta)$, $I_{jmkl} = \int \varphi_{mk}^*(x)\psi_{jl}(x) g(x) dx$ and let

$$z_{\max}(\varphi_{mk}^*, \psi_{jl}) = \arg\max_x [\varphi_{mk}^*(x)\psi_{jl}(x) g(x)]. \tag{9.47}$$

Observe that since $\varphi_{mk}^*(z_{\max}) \ne 0$, we have $I_{jmkl}/A_{kk}^{(m)} = O(1)$. Consider the collection of indices

$$\mathcal{L}_{mjk} = \left\{ l : \; 0 \le l \le 2^j - 1, \; \mathrm{supp}(\varphi_{mk}^*) \cap \mathrm{supp}(\psi_{jl}) \ne \emptyset \right\}.$$

It is easy to see that $\mathcal{L}_{mjk} \subseteq [2^{j-m}(L_\varphi + \delta_b + k) - U_\psi,\; 2^{j-m}(U_\varphi + \delta_b + k) - L_\psi]$, so, for each k, there are $O(2^{j-m})$ terms such that $l \in \mathcal{L}_{mjk}$. Note that $|z_{\max}(\varphi_{mk}^*, \psi_{jl})| \le |z_{\max}(\varphi_{mk}^*)|$ and for each k, there is only finite number of terms such that $|z_{\max}((\varphi_{mk}^*, \psi_{jl})| = |z_{\max}((\varphi_{mk}^*)|$. Indeed, straightforward calculation shows that

$$z_{\max}(\varphi_{mk}^*, \psi_{jl}) = \min[(U_\varphi - \delta_b + k - k_{0m}), 2^{m-j}(U_\psi + l - k_{0j})] \quad \text{if} \quad k_{0m} - k < 0.5\,(U_\varphi + L_\varphi),$$

and

$$z_{\max}(\varphi_{mk}^*, \psi_{jl}) = \max[(L_\varphi + \delta_b + k - k_{0m}), 2^{m-j}(L_\psi + l - k_{0j})] \quad \text{if} \quad k_{0m} - k \ge 0.5\,(U_\varphi + L_\varphi).$$

Hence, $|z_{\max}(\varphi_{mk}^*, \psi_{jl})| = |z_{\max}(\varphi_{mk}^*)|$ if $l \ge 2^{j-m}(U_\varphi - \delta_b + k) - U_\psi$ or $l \le 2^{j-m}(L_\varphi + \delta_b + k) - L_\psi$. Since we also need $l \in \mathcal{L}_{mjk}$, we obtain that $|z_{\max}((\varphi_{mk}^*, \psi_{jl})| = |z_{\max}((\varphi_{mk}^*)|$ if $l \in \mathcal{L}_{mjk}^*$ where

$$\mathcal{L}_{mjk}^* \subseteq [2^{j-m}(L_\varphi + \delta_b + k) - U_\psi, 2^{j-m}(L_\varphi + \delta_b + k) - U_\psi] \cup [2^{j-m}(U_\varphi + \delta_b + k) - U_\psi, 2^{j-m}(U_\varphi + \delta_b + k) - L_\psi]$$

and, thus, $\mathcal{L}_{mjk}^*$ contains at most $2(U_\psi - L_\psi)$ values of l for each k. If $l \in \mathcal{L}_{mjk} \setminus \mathcal{L}_{mjk}^* = \mathcal{L}_{mjk}^c$, then

$$2^{j-m}(L_\varphi + \delta_b + k) - L_\psi < l < 2^{j-m}(U_\varphi - \delta_b + k) - U_\psi. \tag{9.48}$$

Then, by (9.15),

$$\begin{aligned} I_{jmkl} &\sim 2^{-\frac{j-m}{2}} \int \varphi^*(2^{m-j}t + k_{0m} - k)\, \psi(t + k_{0j} - l) 2^{-j\alpha}\, |t|^\alpha \exp(-b|t|^{-\beta} 2^{j\beta}) dt \\ &= O\left(2^{(m-j)/2} 2^{-j\alpha} 2^{-j\beta} 2^{(j-m)\alpha} \exp\left\{ -b 2^{j\beta} |t_{\max}^{(k,l)}|^{-\beta} \right\} \right), \end{aligned}$$

where

$$t_{\max}^{(k,l)} = U_\psi + l - k_{0j} \quad \text{if} \quad k_{0m} - k < (U_\varphi + L_\varphi)/2$$

and

$$t_{\max}^{(k,l)} = L_\psi + l - k_{0j} \quad \text{if} \quad k_{0m} - k \geq (U_\varphi + L_\varphi)/2.$$

Using formula (9.17), we derive that

$$\frac{|I_{jmkl}|}{A_{kk}^{(m)}} = O\left(2^{(j-m)(\beta+1/2)} \exp\left\{-b2^{j\beta}\left[|t_{\max}^{(k,l)}|^{-\beta} - 2^{-(j-m)\beta}|z_{\max}^{(k,k,\delta)}|^{-\beta}\right]\right\}\right), \tag{9.49}$$

where $z_{\max}^{(k,k,\delta)}$ is defined in (9.16).

Denote $h_{jmkl} = |t_{\max}^{(k,l)}| - 2^{(j-m)}|z_{\max}^{(k,k,\delta)}|$ and observe that

$$h_{jmkl} = 2^{(j-m)}(U_\varphi - \delta_b + k) - U_\psi - l \quad \text{if} \quad k_{0m} - k < (U_\varphi + L_\varphi)/2,$$

and

$$h_{jmkl} = l - 2^{(j-m)}(L_\varphi + \delta_b + k) + L_\psi \quad \text{if} \quad k_{0m} - k \geq (U_\varphi + L_\varphi)/2.$$

Comparing the latter formulae with definition of $\mathcal{L}_{mjk}^c$, we derive that, for $l \in \mathcal{L}_{mjk}^c$, $0 < h_{jmkl} < C_h 2^{j-m}$ for every value of k, where $C_h > 0$ is a constant which depends only on the choice of the wavelet basis. Now, for any $0 < x < y$ and $\beta > 0$ one has for some $0 < \xi < y - x$

$$x^{-\beta} - y^{-\beta} = \beta(y-x)(y-\xi)^{-\beta} \geq \beta(y-x)y^{-\beta}.$$

Applying the above inequality with $x = |t_{\max}^{(k,l)}|$ and $y = 2^{(j-m)}|z_{\max}^{(k,k,\delta)}|$, we obtain that

$$|t_{\max}^{(k,l)}|^{-\beta} - 2^{-(j-m)\beta}|z_{\max}^{(k,k,\delta)}|^{-\beta} \geq \beta h_{jmkl} 2^{-(j-m)\beta}|z_{\max}^{(k,k,\delta)}|^{-\beta},$$

and, thus, for $l \in \mathcal{L}_{mjk}^c$, we have

$$(A_{kk}^{(m)})^{-1}\,|I_{jmkl}| = O\left(2^{(j-m)(\beta+1/2)} \exp\left\{-b\beta|z_{\max}^{(k,k,\delta)}|^{-\beta}2^{m\beta}h_{jmkl}\right\}\right). \tag{9.50}$$

Now, it follows from (9.44) that $\Delta_{23} = \Delta_{231} + \Delta_{232}$, where

$$\begin{aligned}
\Delta_{231} &= O\left(\sum_{k \in K_{0m}^\varphi}\left[\sum_{j=m}^{\infty}\sum_{l \in \mathcal{L}_{mjk}^*}(A_{kk}^{(m)})^{-1}\,|I_{jmkl}||b_{jl}|\right]^2\right),\\
\Delta_{232} &= O\left(\sum_{k \in K_{0m}^\varphi}\left[\sum_{j=m}^{\infty}\sum_{l \in \mathcal{L}_{mjk}^c}(A_{kk}^{(m)})^{-1}\,|I_{jmkl}||b_{jl}|\right]^2\right).
\end{aligned}$$

Using the facts thta the set K_{0m}^φ is finite, $|b_{jl}| = O(2^{-js'}$ and $(A_{kk}^{(m)})^{-1}\,|I_{jmkl}| = O(1)$, we derive that, as $m \to \infty$,

$$\Delta_{231} = O\left(2^{-2ms'}\right). \tag{9.51}$$

For Δ_{232}, using (9.50) and taking into account that h_{jmkl} changes by unit incrennents, we obtain

$$\begin{aligned}
\Delta_{232} &= O\left(\sum_{k \in K_{0m}^\varphi}\left[\sum_{j=m}^{\infty}\sum_{h_{jmkl}=0}^{C_h 2^{j-m}} 2^{-js'}2^{(j-m)(\beta+1/2)}\exp\left\{-b\beta|z_{\max}^{(k,k,\delta)}|^{-\beta}2^{m\beta}h_{jmkl}\right\}\right]^2\right)\\
&= o\left(2^{-2ms'}\right).
\end{aligned} \tag{9.52}$$

Finally, combining expressions (9.39), (9.42), (9.43), (9.51) and (9.52), we obtain

$$\Delta_2 = O\left(n^{-1}\, 2^{m\alpha}\ \exp(2b\, 2^{\beta(m+1)}) + 2^{-2ms'}\right). \tag{9.53}$$

To complete the proof of (5.17), set $m = m_0$, where m_0 is defined in (5.18) and combine (9.53) with (9.36), (9.39), (9.42) and (9.43).

Now, we need to show that $\mathbb{E}\|\hat{f}_0^{(m)} - \mathbb{E}\hat{f}_0^{(m)}\|^4 = o(1)$. Note that it follows from (9.37)–(9.40) that $\Delta^* = O\left(\Delta_1^* + \Delta_2^* + \Delta_3^*\right)$ where, similarly to the case of squared difference,

$$\begin{aligned}
\Delta_1^* &= O\left(\|\mathbf{D}^{-1}\|^8 \|\mathbf{Q}^{-1}\|^4\ \mathbb{E}\|\hat{\mathbf{c}} - \mathbf{c}\|^4\right), \\
\Delta_2^* &= O\left(\|\mathbf{D}^{-1}\|^8\ \|\mathbf{B}\|^4 \|\mathbf{Q}^{-1}\|^4\ \mathbb{E}\|\hat{\mathbf{v}} - \mathbf{v}\|^4\right), \\
\Delta_3^* &= O\left(\|\mathbf{D}^{-1}\|^4\ \|\mathbf{Q}^{-1}\|^4\ \|\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^4\right).
\end{aligned}$$

Applying Lemma 9 and using (9.12) and (9.41) with $b = 0$, we obtain $\Delta_1^* = O\left(n^{-2} 2^{2m\alpha}\right) = o(1)$ and $\Delta_2^* = O\left(2^{2m\alpha}\mathbb{E}\|\hat{\mathbf{v}} - \mathbf{v}\|^4\right)$. Also, similarly to (9.44) and (9.46), $\Delta_3^* = O\left(2^{-4ms'}\right)$. To complete the proof of the lemma, recall the definitions of $\hat{\mathbf{v}}$ and $\mathbf{v}$, apply (9.5) with $k \in K_{0m}^*$, and note that, for $k \in K_{0m}^*$, one has $|k - k_{0m}| = O(1)$.

9.6 Proofs of the statements in Section 5: risk of the zero–free part of the wavelet estimator

Proof of Lemma 2 . Let $R = \mathbb{E}\|\hat{f}_{c,m_0} - f_{c,m_0}\|^2 = R_1 + R_2 + R_3$, where

$$R_1 = \sum_{k \in K_{0mc}^{\varphi}} \mathrm{Var}(\hat{a}_{m_0 k}), \quad R_2 = \sum_{j=J}^{\infty} \sum_{k \in K_{0jc}^{\psi}} b_{jk}^2, \quad R_3 = \sum_{j=m_0}^{J-1} \sum_{k \in K_{0jc}^{\psi}} \mathbb{E}(\hat{b}_{jk} - b_{jk})^2.$$

By Lemma 6 we derive that, as $n \to \infty$,

$$\begin{aligned}
R_1 &= O\left(n^{-1}\, 2^{m_0\alpha} \sum_{k \in K_{0m_0c}^{\varphi}} \left[|k - k_{0m_0}|^{-\alpha} \exp(b 2^{m_0\beta} |k - k_{0m_0}|^{-\beta})\right]\right) \\
&= O\left(n^{-1} 2^{m_0(1+\alpha)} \exp(2^{-(\beta+1)} \ln n)\right) = o\left((\ln n)^{-\frac{2s'}{\beta}}\right).
\end{aligned}$$

Using (5.3) and (9.56), we derive that

$$R_2 = O\left(2^{-2Js^*}\right) = O\left((\ln n)^{-\frac{4s^*}{\beta}}\right) = O\left((\ln n)^{-\frac{2s'}{\beta}}\right),$$

since $s^* = s'$ for $1 \le p \le 2$ and $s^* = s \ge (s + 1/2 - 1/p)/2$ for $2 < p \le \infty$ due to $s \ge 1/2$. For R_3, we have

$$\begin{aligned}
R_3 &= \sum_{j=m_0}^{J-1} \sum_{|k-k_{0j}| \le 2^{j-m_0}} b_{jk}^2 + \sum_{j=m_0}^{J-1} \sum_{|k-k_{0j}| > 2^{j-m_0}} \mathrm{Var}(\tilde{b}_{jk}) \\
&= O\left(\sum_{j=m_0}^{J-1} \left[2^{-2js'} (2^{j-m_0})^{1-2/p} + n^{-1}\, 2^j\, 2^{\alpha m_0}\ \exp\left(b 2^{\beta m_0}\right)\right]\right) \\
&= O\left(2^{-m_0(s+2/p-1)} + (\ln n)^{(2+\alpha)/\beta} n^{-1+2^{-(\beta+1)}}\right) = O\left((\ln n)^{-\frac{2s'}{\beta}}\right).
\end{aligned}$$

To complete the proof of the lemma, note that the upper bounds are uniform for $f \in B^s_{p,q}(A)$.

Proof of Lemma 3. Note that

$$R = \mathbb{E}\|\hat{f}_{c,m} - f_c\|^2 = R_1 + R_2 + R_3 + R_4, \tag{9.54}$$

where

$$\begin{aligned}
R_1 &= \sum_{k \in K^{\varphi}_{0mc}} \mathbb{E}(\hat{a}_{mk} - a_{mk})^2, \quad R_2 = \sum_{j=J}^{\infty} \sum_{k \in K^{\psi}_{0jc}} b^2_{jk}, \\
R_3 &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \mathbb{E}\left[(\tilde{b}_{jk} - b_{jk})^2 \, \mathbb{I}(\tilde{b}^2_{jk} > d^2 \, \varrho_n \, 2^{j\alpha} \, |k - k_{0j}|^{-\alpha})\right], \\
R_4 &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} b^2_{jk} \, \mathbb{P}(\tilde{b}^2_{jk} \le d^2 \, \varrho_n \, 2^{j\alpha} \, |k - k_{0j}|^{-\alpha})
\end{aligned}$$

with ϱ_n defined in (9.20). Using Lemma 6, we obtain

$$R_1 = O\left(n^{-1} \, 2^{m\alpha} \sum_{k \in K^{\varphi}_{0mc}} |k - k_{0m}|^{-\alpha}\right) = O\left(n^{-1} \, 2^{m\alpha} (\ln n)^{\mathbb{I}(\alpha=1)}\right) \tag{9.55}$$

since the set K^{φ}_{0mc} contains $O(\ln n)$ terms and the sum $\sum_{k \in K^{\varphi}_{0mc}} |k - k_{0m}|^{-\alpha}$ is uniformly bounded if $\alpha > 1$.

It is well-known (see, e.g., Johnstone (2002), Lemma 19.1) that if $f \in B^s_{p,q}(A)$, then for some constant $c^\star > 0$, dependent on p, q, s and A only, one has

$$\sum_{k=0}^{2^j - 1} b^2_{jk} \le c^\star 2^{-2js^*} \quad \text{with} \quad s^* = \min(s, s'). \tag{9.56}$$

Therefore, an upper bound for R_2 is of the form

$$R_2 = \sum_{j=J}^{\infty} \sum_{k=0}^{2^j - 1} b^2_{jk} = O\left(2^{-2Js'}\right).$$

If $1 \le p \le 2$, then $s^* = s'$ and $R_2 = O\left(n^{-\frac{2s'}{2s'+\alpha}} \, (\ln n)^{\frac{2s'}{2s'+\alpha}}\right)$. If $2 \le p \le \infty$, then $s^* = s$ and, since $s \ge 1/2$, one has $p > (4s - 2\alpha - 2)/(4s^2 - \alpha - 1)$. Hence,

$$2s/(\alpha + 1) > 2s'/(2s' + \alpha),$$

so that, for $1 \le p \le \infty$, one has

$$R_2 = O\left(n^{-\frac{2s'}{2s'+\alpha}} \, (\ln n)^{\frac{2s'}{2s'+\alpha}}\right). \tag{9.57}$$

In oder to obtain an upper bound for R_3 and R_4, note that

$$R_3 \le R_{31} + R_{32}, \quad R_4 \le R_{41} + R_{42}, \tag{9.58}$$

where

$$
\begin{aligned}
R_{31} &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \mathbb{E}\left[(\tilde{b}_{jk} - b_{jk})^2 \, \mathbb{I}((\tilde{b}_{jk} - b_{jk})^2 > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right], \\
R_{32} &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \mathbb{E}\left[(\tilde{b}_{jk} - b_{jk})^2 \, \mathbb{I}(b^2_{jk} > 0.25\, d^2\, \varrho^2_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right], \\
R_{41} &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} b^2_{jk}\, \mathbb{P}((\tilde{b}_{jk} - b_{jk})^2 > 0.25\, d^2\, \varrho^2_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}), \qquad (9.59) \\
R_{42} &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} b^2_{jk}\, \mathbb{I}(b^2_{jk} \le 2.25\ d^2\, \varrho^2_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}).
\end{aligned}
$$

Applying Lemma 10 with $w(\cdot) = \psi(\cdot)$ and $\tau = 0.5d$, we obtain

$$
\mathbb{P}((\tilde{b}_{jk} - b_{jk})^2 > 0.25\, d^2\, \varrho^2_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}) = O\left(n^{-0.5d/C_d}\right),
$$

where C_d is given by (5.20). Hence, by Lemma 6 and inequality $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, for $d > 4C_d$, as $n \to \infty$, we obtain

$$
\begin{aligned}
R_{31} &\le \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \left[\mathbb{E}(\tilde{b}_{jk} - b_{jk})^4 \ \cdot\ \mathbb{P}((\tilde{b}_{jk} - b_{jk})^2 > 0.25\, d^2\, \varrho^2_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right]^{1/2} \\
&= O\left(n^{-\frac{d}{4C_d}} \left[n^{-\frac{3}{2}} 2^{\frac{j(3\alpha+1)}{2}} \sum_{k \in K^{\psi}_{0jc}} |k - k_{0j}|^{-\frac{3\alpha}{2}} + n^{-1} 2^{j\alpha} \sum_{k \in K^{\psi}_{0jc}} |k - k_{0j}|^{-\alpha}\right]\right) \\
&= O\left(n^{-\frac{d}{4C_d}}\right) = o\left(n^{-\frac{2s'}{2s'+\alpha}}\ (\ln n)^{\frac{2s'}{2s'+\alpha}}\right). \qquad (9.60)
\end{aligned}
$$

Similarly, by (9.56),

$$
R_{41} = O\left(n^{-\frac{d}{2C_d}}\right) \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} b^2_{jk} = o(n^{-1}). \qquad (9.61)
$$

Now, consider R_{32} and R_{42}. Note that it follows from Lemma 6 that

$$
\begin{aligned}
R_{32} &= O\left(\sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \left[n^{-1} 2^{j\alpha} |k - k_{0j}|^{-\alpha} \mathbb{I}(b^2_{jk} > 0.5\, d^2\, n^{-1} \ln n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right]\right) \\
&= O\left(\sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \min\left[(\ln n)^{-1} b^2_{jk}, n^{-1} 2^{j\alpha} |k - k_{0j}|^{-\alpha}\right]\right)
\end{aligned}
$$

and, similarly,

$$
R_{42} = O\left(\sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \min\left[b^2_{jk}, n^{-1} \ln n\, 2^{j\alpha} |k - k_{0j}|^{-\alpha}\right]\right).
$$

Hence,

$$R_{32} = O\left((\ln n)^{-1} R_{42}\right) = O\left(R_{42}\right) \tag{9.62}$$

so that one needs to study only R_{42}. Partition R_{42} as $R_{42} = R_{421} + R_{422} + R_{423}$, where

$$\begin{aligned} R_{421} &= \sum_{j=m}^{j_1} \sum_{k \in K^{\psi}_{0jc}} \left[n^{-1} \ln n \, 2^{j\alpha} |k - k_{0j}|^{-\alpha}\right], \quad R_{422} = \sum_{j=j_2}^{J-1} \sum_{k \in K^{\psi}_{0jc}} b^2_{jk}, \\ R_{423} &= \sum_{j=j_1+1}^{j_2-1} \left[\sum_{|k-k_{0j}|>N_j} n^{-1} \ln n \, 2^{j\alpha} |k - k_{0j}|^{-\alpha} + \sum_{|k-k_{0j}|\le N_j} b^2_{jk}\right], \end{aligned}$$

and the values of j_1, j_2 and N_j will be defined later. It is easy to see that, by (9.56),

$$R_{421} = O\left(n^{-1} \ln n \, 2^{j_1 \alpha} (\ln n)^{\mathbb{I}(\alpha=1)}\right), \quad R_{422} = O\left(2^{-2 j_2 s^*}\right), \tag{9.63}$$

$$R_{423} = O\left(\sum_{j=j_1+1}^{j_2-1} \left[n^{-1} \ln n \, 2^{j\alpha} N_j^{1-\alpha} (\ln n)^{\mathbb{I}(\alpha=1)} + 2^{-2js'} N_j^{1-2/p}\right]\right). \tag{9.64}$$

If $\alpha \neq 1$, the two terms in (9.64) are equal to each other when

$$N_j = \left(n^{-1} \ln n 2^{j(2s'+\alpha)}\right)^{1/(\alpha-2/p)},$$

and, for this value of N_j, one has

$$R_{423} = O\left(\sum_{j=j_1+1}^{j_2-1} (n/\ln n)^{\frac{2/p-1}{\alpha-2/p}} 2^{\frac{2j(s'-\alpha s)}{\alpha-2/p}}\right). \tag{9.65}$$

Therefore, R_{423} behave differently when $\alpha s \geq s'$ and $\alpha s < s'$, and we consider those cases separately.

First, consider the case when $\alpha s = s'$. Then

$$R_{423} = O\left((j_2 - j_1)(n/\ln n)^{\frac{2/p-1}{\alpha-2/p}}\right) = O\left((\ln n/n)^{\frac{2s'}{2s'+\alpha}} \ln n\right) \quad \text{if} \quad \alpha s = s'.$$

If $\alpha > 1$, $\alpha s > s'$, choose j_1 and j_2 such that

$$2^{j_1} = (n/\ln n)^{\frac{1}{2s'+\alpha}}, \quad 2^{j_2} = (n/\ln n)^{\frac{s'}{s^*(2s'+\alpha)}}.$$

Note that if $1 \leq p \leq 2$, one has $s^* = s \geq s'$, so that $j_2 \leq j_1$ and $R_{423} = 0$. If $2 < p \leq \infty$, then $j_2 > j_1$. Also, it follows from (9.63) and (9.65) that $R_{423} = O\left(n^{\frac{2/p-1}{\alpha-2/p}} (\ln n)^{\frac{1-\alpha}{\alpha-2/p}} 2^{\frac{2j_1(s'-\alpha s)}{\alpha-2/p}}\right)$. Hence, $R_{421} = O\left((n/\ln n)^{-\frac{2s'}{2s'+\alpha}}\right)$, $R_{422} = O\left((n/\ln n)^{-\frac{2s'}{2s'+\alpha}}\right)$ and $R_{423} = O\left((n/\ln n)^{-\frac{2s'}{2s'+\alpha}}\right)$, so that

$$R_{42} = O\left((n/\ln n)^{-\frac{2s'}{2s'+\alpha}}\right) \quad \text{if} \quad \alpha s \geq s', \, \alpha > 1. \tag{9.66}$$

Similarly, if $\alpha > 1$, $\alpha s < s'$, choose j_1 and j_2 such that

$$2^{j_1} = (n/\ln n)^{\frac{1}{\alpha(2s+1)}}, \quad 2^{j_2} = (n/\ln n)^{\frac{1}{2s+1}}.$$

In this case, $R_{423} = O\left(n^{\frac{2/p-1}{\alpha-2/p}}(\ln n)^{\frac{1-\alpha}{\alpha-2/p}}2^{\frac{2j_2(s'-\alpha s)}{\alpha-2/p}}\right)$, and direct calculations yield $R_{421} = O\left((n/\ln n)^{-\frac{2s}{2s+1}}\right)$, $R_{422} = O\left((n/\ln n)^{-\frac{2s}{2s+1}}\right)$ and $R_{423} = O\left((n/\ln n)^{-\frac{2s}{2s+1}}\right)$, so that

$$R_{42} = O\left((n/\ln n)^{-\frac{2s}{2s+1}}\right) \quad \text{if} \quad \alpha s < s',\ \alpha > 1. \tag{9.67}$$

Finally, if $\alpha = 1$, set $j_1 = j_2$ such that

$$2^{j_1} = (n/\ln^2 n)^{\frac{1}{2s^*+1}}$$

and obtain

$$R_{42} = O\left(n^{-\frac{2s^*}{2s^*+1}}\ (\ln n)^{\frac{4s^*-1}{2s^*+1}}\right) \quad \text{if} \quad \alpha = 1. \tag{9.68}$$

Now, to complete the proof of (5.21), one just need to combine (9.54), (9.55), (9.57), (9.60), (9.61), (9.62) and (9.66)–(9.68), and to note that all upper bounds are uniform for $f \in B^s_{p,q}(A)$.

In order to prove (5.22), note that

$$R^* = \mathbb{E}\|\hat{f}_{c,m} - f_c\|^4 \le R^*_1 + R^*_2 + R^*_3,$$

where

$$\begin{aligned}
R^*_1 &= O\left(\mathbb{E}\|\sum_{k\in K^{\varphi}_{0mc}}(\hat{a}_{mk} - a_{mk})\varphi_{mk}(x)\|^4\right), \quad R^*_2 = O\left(\|\sum_{j=m}^{\infty}\sum_{k\in K^{\psi}_{0jc}} b_{jk}\psi_{jk}(x)\|^4\right),\\
R^*_3 &= O\left(\mathbb{E}\left[\sum_{j=m}^{J-1}\sum_{k\in K^{\psi}_{0jc}}(\tilde{b}_{jk} - b_{jk})^2\ \mathbb{I}(\tilde{b}^2_{jk} > d^2\ \varrho_n\ 2^{j\alpha}\,|k - k_{0j}|^{-\alpha})\right]^2\right).
\end{aligned}$$

Observe that, by Lemma 6, since $2^{m(\alpha+1)} = o(n/\ln n)$, as $n \to \infty$,

$$\begin{aligned}
R^*_1 &= O\left(2^m \sum_{k\in K^{\varphi}_{0mc}}\mathbb{E}(\hat{a}_{mk} - a_{mk})^4\right) = O\left(n^{-3}\,2^{m(3\alpha+2)} + n^{-2}\,2^{m(2\alpha+1)}\right)\\
&= O\left(n^{-2}\,2^{m(2\alpha+1)}\right) = o(1).
\end{aligned}$$

For R^*_2, by (9.56), we have

$$R^*_2 = O\left(\left[\sum_{j=m}^{\infty}\sum_{k\in K^{\psi}_{0jc}} b^2_{jk}\right]^2\right) = O\left(\left[2^{-2ms'}\right]^2\right) = o(1).$$

Finally, similarly to (9.59), partition R^*_3 as $R^*_3 = R^*_{31} + R^*_{32}$ with R^*_{31} and R^*_{32} corresponding to $\mathbb{I}(|\tilde{b}_{jk} - b_{jk}|^2 > 0.25\,d^2\ \varrho_n\ 2^{j\alpha}\,|k - k_{0j}|^{-\alpha})$ and $\mathbb{I}(b^2_{jk} > 0.25\,d^2\ \varrho_n\ 2^{j\alpha}\,|k - k_{0j}|^{-\alpha})$, respectively. For R^*_{31}, applying Lemmas 6 and 10 with $w = \psi$ and C_d given by (5.20), and also noting that

$\sum_{k\in K^{\psi}_{0jc}} |k-k_{0j}|^{-l} = O(1)$ for $l>1$, we derive

$$
\begin{aligned}
R^*_{31} &= O\left(n \sum_{j=m}^{J-1} \sum_{k\in K^{\psi}_{0jc}} \mathbb{E}\left[|\tilde{b}_{jk} - b_{jk}|^4 \, \mathbb{I}(|\tilde{b}_{jk} - b_{jk}|^2 > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k-k_{0j}|^{-\alpha}) \right] \right) \\
&= O\left(n \sum_{j=m}^{J-1} \sum_{k\in K^{\psi}_{0jc}} \left[\mathbb{E}|\tilde{b}_{jk} - b_{jk}|^6 \right]^{2/3} \left[\mathbb{P}(|\tilde{b}_{jk} - b_{jk}|^2 > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k-k_{0j}|^{-\alpha}) \right]^{1/3} \right) \\
&= O\left(n \sum_{j=m}^{J-1} n^{-d/(3C_d)} \left[n^{-10/3}\, 2^{j(10\alpha+4)/3} + n^{-8/3}\, 2^{j(8\alpha+2)/3} + n^{-2}\, 2^{2j\alpha} \right] \right) \\
&= o\left(n^{1-d/(3C_d)} \right) = o(1), \quad n\to\infty,
\end{aligned}
$$

since $d > 3C_d$. For R^*_{32}, using Lemma 6 and (9.56), we derive that

$$
\begin{aligned}
R^*_{32} &= O\left(n \sum_{j=m}^{J-1} \sum_{k\in K^{\psi}_{0jc}} \mathbb{E}\left[|\tilde{b}_{jk} - b_{jk}|^4 \, \mathbb{I}(b^2_{jk} > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k-k_{0j}|^{-\alpha}) \right] \right) \\
&= O\left(n \sum_{j=m}^{J-1} \sum_{k\in K^{\psi}_{0jc}} [2^j \ \ln^{-3} n\, b^6_{jk} + \ln^{-2} n\, b^4_{jk}] \right) = o\left(n \sum_{j=m}^{J-1} [2^{j(1-6s')} + 2^{-4js'}] \right)
\end{aligned}
$$

since $n^{-1}\, 2^{j\alpha} |k-k_{0j}|^{-\alpha} < 0.25\, b^2_{jk}/(d^2 \ln n)$. Note that $m \geq m_0$ implies $2^m \geq n^{1/(2s'+\alpha)}$, so that

$$
R^*_{32} = o\left(n^{-\frac{6s'-1}{2s'+\alpha}} + n^{-\frac{4s'}{2s'+\alpha}} \right) = o(1),
$$

which completes the proof of the lemma.

9.7 Proof of the minimax upper bounds for the risk in Section 6

Proof of Theorem 2. Since $\hat{m} = m_0$ for $b>0$, the validity of Theorem 2 for $b>0$ follows directly from Lemma 2. For $b=0$, observe that

$$
\Delta = \mathbb{E}[\|\hat{f}_m - f\|^2 = \sum_{m=m_1}^{m_0} \mathbb{E}[\|\hat{f}_m - f\|^2 \ \ \mathbb{I}(\hat{m} = m \leq m_0)] + \mathbb{E}[\|\hat{f}_m - f\|^2 \ \ \mathbb{I}(\hat{m} = m > m_0)] \equiv \Delta_1 + \Delta_2
$$

and consider terms Δ_1 and Δ_2 separately.

Denote

$$
R(n) = \begin{cases} O\left(n^{-\frac{2s}{2s+1}} \ (\ln n)^{\mu_1} \right) & \text{if} \quad b=0, \ \alpha s < s', \\ O\left(n^{-\frac{2s'}{2s'+\alpha}} \ (\ln n)^{\mu_2} \right) & \text{if} \quad b=0, \ \alpha s \geq s', \end{cases} \tag{9.69}
$$

and note that for any $m \geq m_1$

$$
\mathbb{E}\|\hat{f}_m - f\|^2 \leq 2[\,\mathbb{E}\|\hat{f}_{m_0} - f\|^2 + \mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x\in\Xi_m)\|^2 + \mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x\in\Xi^c_m)\|^2]
$$

where m_1 is defined in (5.3) and set Ξ_m is defined in (6.1). By Lemmas 1 and 3, we obtain

$$\mathbb{E}\|\hat{f}_{m_0} - f\|^2 = O\left(n^{-\frac{2s'}{2s'+\alpha}} + R(n)\right).$$

If $\hat{m} = m \le m_0$, then by definition of $\hat{m}$, we derive that

$$\mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Xi_m)\|^2 \le \lambda^2\, 2^{m_0\alpha}\, n^{-1} \ln n = O\left(n^{-\frac{2s'}{2s'+\alpha}}\right).$$

Now, recall that Ξ_m is defined in such a way that $\mathrm{supp}(f_{0,m}) \in \Xi_m$ for any m and $\Xi_{j1} \subset \Xi_{j2}$ for $j_1 > j_2$, so that for $m \le m_0$ one has

$$\begin{aligned}
\mathbb{E}\|(\hat{f}_m - f)\mathbb{I}(x \in \Xi_m^c)\|^2 &= \mathbb{E}\|(\hat{f}_{0,m} + \hat{f}_{c,m} - f_{0,m} - f_{c,m})\mathbb{I}(x \in \Xi_m^c)\|^2 \\
&= \mathbb{E}\|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(x \in \Xi_m^c)\|^2 \le \mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^2 = O\left(R(n)\right)
\end{aligned}$$

as $n \to \infty$. Noting that

$$\mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Xi_m^c)\|^2 \le 2\left[\mathbb{E}\|(\hat{f}_m - f)\mathbb{I}(x \in \Xi_m^c)\|^2 + \mathbb{E}\|(\hat{f}_{m_0} - f)\mathbb{I}(x \in \Xi_m^c)\|^2\right]$$

and combining all formulae above, we obtain that $\Delta_1 = O\left(R(n)\right)$ as $n \to \infty$.

By Lemmas 1 and 3, one has $\mathbb{E}\|\hat{f}_{0,m} - f_{0,m}\|^4 = o\,(1)$ and $\mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^4 = o\,(1)$. Then, Lemma 4 yields

$$\Delta_2 \;\le\; \sqrt{\mathbb{E}[\|\hat{f}_m - f\|^4}\,\sqrt{\mathbb{P}(\hat{m} = m > m_0)} = O\left(n^{-\frac{\lambda}{2C_\lambda}} + n^{\frac{1}{2(\alpha+1)} - \frac{d}{4C_d}}\right) = O(n^{-1})$$

provided $\lambda \ge \max\left(C_{\lambda 1}, C_{\lambda 2}, 2C_\lambda\right)$ and $d > 2(\alpha+1)^{-1}(2\alpha+3)C_d$, which completes the proof of Theorem 2.

9.8 Proofs of the statements in Section 7

Proof of Theorem 3 is based on the following lemma.

Lemma 13 *Let Assumption A hold with $b = 0$ and $0 < \alpha < 1$. Then,*

$$\begin{aligned}
Var(\tilde{b}_{jk}) &= O\left(n^{-1} 2^{j\alpha}\, \min(1, |k - k_{0j}|^{-\alpha})\right), \\
\mathbb{E}|\tilde{b}_{jk} - b_{jk}|^{\frac{\alpha+3}{\alpha+1}} &= O\left(n^{-\frac{2}{\alpha+1}} 2^{j\frac{(\alpha+3)}{2(\alpha+1)}} + n^{-\frac{\alpha+3}{2(\alpha+2)}} 2^j\right)
\end{aligned}$$

Proof of Lemma 13. Proof of the first statement is very similar to the proof of validity of formula (9.6). Proof of the second statement is based on Lemma 3.1. in Chesneau (2007a) which states that whenever $\int [g(x)]^{1-\nu} dx < \infty$ for some $\nu > 2$, one has

$$\mathbb{E}|\tilde{b}_{jk} - b_{jk}|^\nu = O\left(n^{1-\nu} \int |\psi_{jk}(x)|^\nu\, [g(x)]^{1-\nu} dx + n^{-\nu/2} \int \psi_{jk}^2(x)\, [g(x)]^{-\nu/2} dx\right). \tag{9.70}$$

To complete the proof, note that for $\nu = 1 + 2/(\alpha+1) > 2$ one has $\int [g(x)]^{1-\nu} dx < \infty$ and apply (9.70).

Proof of Theorem 3. Proof of this statement is similar to the proof of Lemma 3. Indeed, similarly to the proof of Lemma 3, partition the risk as $R = \mathbb{E}\|\hat{f} - f\|^2 = R_1 + R_2 + R_3 + R_4$ where, similarly to the proof of Lemma 3,

$$
\begin{aligned}
R_1 &= \sum_{k=0}^{2^{m_1}-1} \mathbb{E}(\hat{a}_{m_1 k} - a_{m_1 k})^2, \quad R_2 = \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} b_{jk}^2, \\
R_3 &= \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[(\tilde{b}_{jk} - b_{jk})^2 \, \mathbb{I}(\tilde{b}_{jk}^2 > d^2 \, \varrho_n \, 2^{j\alpha} \, |k - k_{0j}|^{-\alpha}) \right], \\
R_4 &= \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} b_{jk}^2 \, \mathbb{P}(\tilde{b}_{jk}^2 \le d^2 \, \varrho_n \, 2^{j\alpha} \, |k - k_{0j}|^{-\alpha})
\end{aligned}
$$

with ϱ_n defined in (9.20). Since $1/g$ is integrable and m_1 in (5.3) is finite, it is easy to show that $R_1 = O\left(n^{-1}\right)$. Also, same as before, $R_2 = O\left(2^{-2Js^*}\right)$. If $p > 2$, then $\alpha + 1 < 2s + 1$ since $s \ge \max(1/2, 1/p)$ and $\alpha < 1$, so that $R_2 = O\left(n^{-2s/(2s+1)}\right)$. If $1 \le p \le 2$, then $s^* = s'$ and $2s'/(1+\alpha) > \max\left\{2s'/(2s'+\alpha), 2s/(2s+1)\right\}$, so that

$$
R_2 = O\left(\max\left\{ n^{-2s/(2s+1)}, n^{-2s'/(2s'+\alpha)} \right\} \right).
$$

Now, similarly to the proof of Lemma 3, partition R_3 and R_4 as $R_3 \le R_{31} + R_{32}$ and $R_4 \le R_{41} + R_{42}$. Using Lemma 13, as $n \to \infty$, obtain upper bounds

$$
\begin{aligned}
R_{31} &\le \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \left[\mathbb{P}((\tilde{b}_{jk} - b_{jk})^2 > 0.25 \, d^2 \, \varrho_n^2 \, 2^{j\alpha} \, |k - k_{0j}|^{-\alpha}) \right]^{1-2/\nu} \left[\mathbb{E}|\tilde{b}_{jk} - b_{jk}|^{\nu} \right]^{2/\nu} \\
&= O\left(\sum_{j=0}^{J-1} 2^j \, n^{-d(1-2/\nu)/(2C_d)} \, \left[n^{1-\nu} 2^{j\nu/2} + n^{-\nu/2} \, 2^j \right]^{2/\nu} \right) \\
&= O\left(2^j \, n^{-d(1-2/\nu)/(2C_d)} \, \left[n^{-\nu/2} \, 2^J \right]^{2/\nu} \right) = O\left(n^{-1}\right),
\end{aligned}
$$

provided (7.2) holds, and also

$$
R_{41} \; = \; O\left(n^{-\frac{d}{2C_d}} \right) \sum_{j=m}^{J-1} \sum_{k \in K_{0jc}^{\psi}} b_{jk}^2 = o(n^{-1}).
$$

Now, same as before, $R_{32} = O\left((\ln n)^{-1} R_{42}\right) = O\left(R_{42}\right)$, so that we need to construct upper bounds for R_{42} only. Partition R_{42} as $R_{42} = R_{421} + R_{422} + R_{423}$ where

$$
\begin{aligned}
R_{421} &= \sum_{j=0}^{j_1} \sum_{k=0}^{2^j-1} \left[n^{-1} \, \ln n \, 2^{j\alpha} |k - k_{0j}|^{-\alpha} \right], \quad R_{422} = \sum_{j=j_2}^{J-1} \sum_{k=0}^{2^j-1} b_{jk}^2, \\
R_{423} &= \sum_{j=j_1+1}^{j_2-1} \left\{ \sum_{|k-k_{0j}| > N_j} |b_{jk}|^p \left[n^{-1} \, \ln n \, 2^{j\alpha} N_j^{-\alpha} \right]^{1-p/2} + \sum_{|k-k_{0j}| \le N_j} n^{-1} \, \ln n \, 2^{j\alpha} N_j^{1-\alpha} \right\},
\end{aligned}
$$

and the values of j_1, j_2 and N_j will be defined later. It is easy to see that, same as before, $R_{421} = O\left(n^{-1} \, \ln n \, 2^{j_1 \alpha}\right)$ and $R_{422} = O\left(2^{-2j_2 s^*}\right)$. For R_{423} we can write the following expression

$$
R_{423} \; = \; \sum_{j=j_1+1}^{j_2-1} \left[2^{-js'p} \left(\frac{\ln n}{n} \, 2^{j\alpha} N_j^{-\alpha} \right)^{1-p/2} + \frac{\ln n}{n} \, 2^{j\alpha} N_j^{1-\alpha} \right].
$$

If $p \geq 2$, we choose $j_1 = j_2$ such that $2^{j_1} = (\ln n/n)^{1/(2s+1)}$ and obtain $R_{42} = O\left((\ln n/n)^{2s/(2s+1)}\right)$. If $1 \leq p < 2$, we choose N_j which equalize the two terms in R_{423} and obtain, similarly to (9.65),

$$R_{423} = \begin{cases} O\left((n/\ln n)^{\frac{2/p-1}{\alpha-2/p}} 2^{\frac{2j_2(s'-\alpha s)}{\alpha-2/p}}\right) & \text{if} \quad \alpha s < s' \\ O\left((n/\ln n)^{\frac{2/p-1}{\alpha-2/p}} 2^{\frac{2j_1(s'-\alpha s)}{\alpha-2/p}}\right) & \text{if} \quad \alpha s > s' \\ O\left((j_2 - j_1)\,(n/\ln n)^{\frac{2/p-1}{\alpha-2/p}}\right) & \text{if} \quad \alpha s = s' \end{cases} \tag{9.71}$$

If $\alpha s < s'$, then choose

$$2^{j_1} = (n/\ln n)^{\frac{1}{2s+1}}, \quad 2^{j_2} = (n/\ln n)^{\frac{s}{s'(2s+1)}},$$

so that $j_1 < j_2$. Direct calculations show that in this case

$$R_{42} = O\left((n/\ln n)^{\zeta_1}\right) \quad \textit{with} \quad \zeta_1 = \frac{2/p-1}{2/p-\alpha} + \frac{2(s'-\alpha s)}{(2s+1)(2/p-\alpha)} = \frac{2s}{2s+1}$$

and $R_{42} = O\left((\ln n/n)^{2s/(2s+1)}\right)$. If $\alpha s > s'$, then set

$$2^{j_1} = (n/\ln n)^{\frac{\alpha}{2s'+\alpha}}, \quad 2^{j_2} = (n/\ln n)^{\frac{1}{2s'+\alpha}},$$

so that again $j_1 < j_2$. Here we have

$$R_{42} = O\left((n/\ln n)^{\zeta_2}\right) \quad \textit{with} \quad \zeta_2 = \frac{2/p-1}{2/p-\alpha} - \frac{2(\alpha s - s')}{(2s'+\alpha)(2/p-\alpha)} = \frac{2s'}{2s'+\alpha}$$

and $R_{42} = O\left((\ln n/n)^{2s'/(2s'+\alpha)}\right)$. If $\alpha s = s'$, then note that $j_2 - j_1 = O(\ln n)$, so that $R_{42} = O\left((\ln n/n)^{2s'/(2s'+\alpha)}\right) = O\left((\ln n/n)^{2s/(2s+1)}\right)$. Now, to complete the proof, just combine the expressions for R_1, R_2, R_{31}, R_{41}, R_{32} and R_{42}.

References

[1] Amato, U., Antoniadis, A. and Pensky, M. (2006). Wavelet kernel penalized estimation for non-equispaced design regression. *Statistics and Computing*, **16**, 37–55.

[2] Antoniadis, A. and Pham, D.T. (1998). Wavelet regression for random or irregular design. *Computational Statistics and Data Analysis*, **28**, 353–369.

[3] Brown, L.D., Cai, T.T., Low, M.G. and Zhang, C.H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Annals of Statistics*, **30**, 688–707.

[4] Cai, T.T. and Brown, L.D. (1998). Wavelet shrinkage for nonequispaced samples. *Annals of Statistics*, **26**, 1783–1799.

[5] Chesneau, C. (2007a). Regression in random design: a minimax study. *Statistics and Probability Letters*, **77**, 40–53.

[6] Chesneau, C. (2007b). Wavelet block thresholding for samples with random design: A minimax approach under the L_p risk. *Electronic Journal of Statistics*, **1**, 331–346.

[7] Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.

[8] Dingle, R.B. (1973). *Asymptotic Expansions: Their Derivation and Interpretation*. London: Academic Press.

[9] Gaïffas, S. (2005). Convergence rates for pointwise curve estimation with a degenerate design. *Mathematical Methods of Statistics*, **14**, 1–27.

[10] Gaïffas, S. (2006). Sharp estimation in sup norm with random design. *Statistics and Probability Letters*, **77**, 782–794.

[11] Gaïffas, S. (2007). On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM: Probability and Statistics*, **11**, 344–364.

[12] Gaïffas, S. (2009). Uniform estimation of a signal based on inhomogeneous data. *Statistica Sinica*, **19**, 427–447.

[13] Hall, P. and Turlach, B.A. (1997). Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Annals of Statistics*, **25**, 1912–1925.

[14] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*. Lecture Notes in Statistics **129**, New York: Springer-Verlag.

[15] Hoffmann, M. and Reiss, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *Annals of Statistics*, **36**, 310--336.

[16] Johnstone, I.M. (1994). Minimax Bayes, asymptotic minimax and sparse wavelet priors. In *Statistical Decision Theory and Related Topics*, (Ed. S.S. Gupta and J.O. Berger), Vol. **V**, pp. 303–326, New York: Springer-Verlag.

[17] Johnstone, I.M. (2002). *Function Estimation and Gaussian Sequence Models*. Unpublished Monograph. (http://www-stat.stanford.edu/∼imj/).

[18] Johnstone, I.M., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society*, Series B, **66**, 547–573 (with discussion, 627–657).

[19] Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli*, **10**, 1053–1105.

[20] Kohler, M. (2003). Nonlinear orthogonal series estimation for random design regression. *Journal of Statistical Planning and Inference*, **115**, 491–520.

[21] Korostelev, A.P. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction.* Lecture Notes in Statistics, **82**, New York: Springer-Verlag.

[22] Kovac, A. and Silverman, B.W. (2000). Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *Journal of the American Statistical Association*, **95**, 172–183.

[23] Kulik, R. and Raimondo, M. (2009). Wavelet regression in random design with heteroscedastic dependent errors. *Annals of Statistics*, **37**, 3396–3430.

[24] Lepski, O.V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Appications*, **35**, 454–466.

[25] Lepski, O.V. (1991). Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimators. *Theory of Probability and its Appications*, **36**, 682–697.

[26] Lepski, O.V., Mammen, E. and Spokoiny, V.G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors *Annals of Statistics*, **25**, 929–947.

[27] Lepski, O. and Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics*, **25**, 2512–2546.

[28] Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. 2nd Edition, San Diego: Academic Press.

[29] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press.

[30] Pensky, M. and Sapatinas, T. (2009). Functional deconvolution in a periodic case: uniform case. *Annals of Statistics*, **37**, 73–104.

[31] Pensky, M. and Sapatinas, T. (2010). On convergence rates equivalency and sampling strategies in functional deconvolution models. *Annals of Statistics*, **38**, 1793–1844.

[32] Pensky, M. and Vidakovic, B. (2001). On non-equally spaced wavelet regression. *Annals of the Institute of Statistical Mathematics*, **53**, 681–690.

[33] Sardy, S., Percival, D.B., Bruce, A.G., Gao, H.-Y. and Stuelzle, W. (1999). Wavelet shrinkage for unequally spaced data. *Statistics and Computing*, **9**, 65–75.

[34] Tsybakov, A.B. (2008). *Introduction to Nonparametric Estimation*. New York: Springer-Verlag.

[35] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM: Philadelphia.

[36] Zhang, S., Wong, M.-Y. and Zheng, Z. (2002). Wavelet threshold estimation of a regression function with random design. *Journal of Multivariate Analysis*, **80**, 256–284.